Cis-regulatory chromatin loops arise before TADs and gene activation and are independent of cell fate during early *Drosophila* development

Sergio Martin Espinola[1*], Markus Götz[1*], Maelle Bellec[2], Olivier Messina[1,2], Jean-Bernard Fiche[1], Christophe Houbron[1], Matthieu Dejean[2], Ingolf Reim[3], Andrés M. Cardozo Gizzi[4], Mounia Lagha[2#], Marcelo Nollmann[1#]

[1] *Centre de Biologie Structurale, CNRS UMR 5048, INSERM U1054, Univ Montpellier, Montpellier, France*

[2] *IGMM, CNRS, Univ Montpellier, Montpellier, France*

[3] *Department of Biology, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany.*

[4] *Centro de Investigación en Medicina Traslacional Severo Amuchastegui (CIMETSA), Instituto Universitario de Ciencias Biomédicas de Córdoba (IUCBC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Naciones Unidas 420, X5016KEJ, Córdoba, Argentina*

** Co-first authors*

 # Correspondence: marcelo.nollmann@cbs.cnrs.fr, mounia.lagha@igmm.cnrs.fr

Abstract

Acquisition of cell fate is thought to rely on the specific interaction of remote *cis*-regulatory modules (CRMs), e.g. enhancers, and target promoters. However, the precise interplay between chromatin structure and gene expression is still unclear, particularly within multicellular developing organisms. Here we employ Hi-M, a single-cell spatial genomics approach, to detect CRM-promoter looping interactions within topologically associating domains (TADs) during early *Drosophila* development. By comparing *cis*-regulatory loops in alternate cell types, we show that physical proximity does not necessarily instruct transcriptional states. Moreover, multi-way analyses reveal multiple CRMs spatially coalesce to form hubs. Loops and CRM hubs are established early during development, prior to the emergence of TADs. Moreover, CRM hubs are formed, in part, via the action of the pioneer transcription factor Zelda and precede transcriptional activation. Our approach provides insight into the role of CRM-promoter interactions in defining transcriptional states, as well as distinct cell types.

Introduction

Chromosomes are organized at different levels —nucleosomes, chromatin loops, TADs and chromosome territories— and each of these layers contributes to the regulation of transcription [1,2]. Particularly, loops between enhancers (E) and promoters (P) are critical for the precise regulation of transcriptional activation [3–7]. In addition, organization of chromosomes into TADs plays a role in transcriptional regulation [8], primarily by facilitating communication between enhancers and promoters through E-P loops within a TAD and restricting contacts from enhancers of neighboring TADs [5,9–13]. However, the interplay between formation of E-P loops, emergence of TADs, and transcriptional output is still poorly understood [14].

Tissue-specific enhancers have been shown to be in close proximity [15] to their cognate promoters indicating that E-P contacts are needed for precise gene regulation [16–19]. Indeed, introduction of ectopically enforced E-P contacts can lead to transcriptional activation of a reporter gene during *Drosophila* development [20]. In some cases, enhancers can increase transcriptional output by modulating transcriptional bursting [6,21–24]. However, in other cases, E-P contacts seem to be dissociated from gene activation[25,26], suggesting that an enhancer may not necessarily need to be in continuous physical contact with a promoter to influence transcription. The mechanisms by which E-P contacts may regulate transcription are currently under intense debate [14,27–29].

Promoters can contact several distant enhancers [16–18], raising the possibility that more than one enhancer may contact a promoter at any given time. More recently, use of multi-way 3C and 4C methods showed that, indeed, enhancers can cluster together to form enhancer hubs [30–33]. This is supported by evidence of nuclear microenvironments containing multiple enhancers and clusters of transcription factors (TFs)[34–40]. This model is consistent with multi-way interactions between distal enhancers to regulate promoter activity of single or multiple genes by sharing resources. Whether and how formation of multi-way interactions may be related to the emergence of TADs during development [41,42] is still an open question.

To shed light onto these questions, we investigated the interplay between transcriptional state and physical proximity between promoters and large sets of CRMs (e.g. enhancers, silencers) during the awakening of the zygotic genome in early *Drosophila* embryos. During the first hours of development, *Drosophila* embryos offer an ideal biological context to decipher how CRMs are employed to establish precise spatiotemporal patterns of gene expression. Decades of genetic and genomic studies have characterized CRMs at a large scale and their usage to interpret morphogen gradients [43–45]. In particular, the pioneering activity of factors such as Zelda (Zld) establishes early accessibility of CRMs (reviewed in [46]) and is involved in the emergence of TAD organization [41,47].

Here, we used Hi-M, an imaging-based technology enabling the detection of chromatin organization and transcriptional status in intact embryos [48,49]. This technology allowed us to visualize where and when interactions between CRMs occur and investigate their impact on transcriptional states. We first used Hi-M to detect intra-TAD chromatin loops in *Drosophila* embryos. We show that the majority of these loops involve CRMs. In fact, we identified not only E-P loops but multiple CRM contacts (E-P, P-P and E-E) co-interacting locally in single nuclei and referred to as CRM hubs. Unexpectedly, these contacts were not found to be specific to transcriptionally active nuclei. Hence, tissues with different cell fates, exhibit similar CRMs contacts and E-P loops. Moreover, networks of CRM loops are established at early stages, prior to the emergence of TADs and before transcriptional activation. Finally, the pioneer factor Zld is required for the establishment of subsets of CRM hubs.

Results

High-resolution Hi-M reveals preferential interactions between *cis*-regulatory modules

Functional characterization of specific chromatin loops between CRMs within TADs (Fig. 1a) requires the development of technologies adapted for the simultaneous detection of such looping interactions and of transcriptional output. Recently, we and others established a new family of imaging-based methods able to retrieve chromatin architecture and transcriptional status simultaneously in single cells (Hi-M and ORCA) [48–50]. Hi-M relies on the labeling and imaging of the expression pattern of genes by direct detection of transcripts via RNA-FISH, followed by the sequential imaging of tens of distinct DNA loci by oligopaint-FISH [51] in intact *Drosophila* embryos [48,49]. First, we tested whether conventional Hi-M was able to detect intra-TAD chromatin loops in two genomic regions harboring early developmental genes expressed at different timings and regions of the embryo (*dorsocross* (*doc*)- and *snail* (*sna*)-TADs).

The doc-TAD contains a family of three genes, the *dorsocross* genes *doc1*, *doc2*, and *doc3* encoding functionally redundant T-box transcription factors essential for the development of the amnioserosa and cardiogenesis [52]. These genes display similar expression patterns, particularly during early stages of embryogenesis, in the blastoderm embryo (nuclear cycle, nc 11 to 14), which will be the focus of this study (Extended Data Fig. 1a-b). In early embryos, the doc-TAD is flanked by insulator binding sites (e.g. CP190), and displays extensive H3K27me3 marks as well as several prominent Zelda peaks (Fig. 1b) [53–56]. At nc14, the Hi-M contact probability map of this genomic region displays two clear TADs, similar to those detected by Hi-C (TAD1, and doc-TAD, Fig. 1b and Extended Data Fig. 1a) [47]. Inspection of ATAC-seq [57], H3K4me3, H3K4me1, and H3K27ac profiles[58] as well as of enhancer databases [59] revealed that the doc-TAD contains several putative CRMs, including four potential enhancers (CRM_a, CRM_b, CRM_c and CRM_d) for the three *doc* promoters (Fig. 1c, and Supplementary Table 1,) [53,54,59,60]. We note that only CRM_a displayed exhaustive binding by several chromatin insulators (Extended Data Fig. 1b). Conventional Hi-M/Hi-C did not exhibit clear specific looping interactions within the doc-TAD, most likely due to insufficient genomic resolution and coverage (Fig. 1b).

To overcome these limitations and probe communications between CRMs and promoters within TADs in an unbiased manner, we improved the genomic resolution and coverage of Hi-M by 3-fold (from ~8-10 kb to ~3 kb) and painted the entire doc-TAD with contiguous barcodes (Extended Data Fig. 1a, c-i), particularly targeting promoters and predicted CRMs (Fig. 1c and Extended Data Fig. 1a). We first focused on enhancers already validated by transgenic assays (CRM_{b-d}) (Supplementary Table 1). The three *doc* genes within the doc-TAD exhibit a shared spatiotemporal profile of expression in late nc14 (Extended Data Fig. 2a). The frequency of activation, estimated by the number of alleles transcribing per nucleus [21,61], was elevated for both *doc1* and *doc2* (~90%) (Extended Data Fig. 2b). In addition, *doc* genes displayed a high degree of co-activation (~70%, Extended Data Fig. 2c). Thus, we hypothesized that multiple putative CRMs are likely to contact *doc* promoters to regulate their common expression pattern.

To test this hypothesis, we selected nuclei displaying at least one nascent mRNA-FISH *doc1* spot and built a high-resolution Hi-M contact map containing only these nuclei (Fig. 1d-f). Remarkably, the improvement in genomic coverage in Hi-M now enabled the detection of specific looping interactions between genetic elements within the doc-TAD in intact embryos (Fig. 1f and Extended Data Fig. 1g). The strongest contacts represented in all cases interactions between CRMs (Fig. 1f, yellow arrows), but there was a considerable inter-nuclear variation (see single nucleus snapshots in Fig. 1f). Contact frequencies did not vary considerably when only nuclei displaying the strongest *doc1* RNA-FISH signals were used to construct the matrix (Extended Data Fig. 2d), suggesting that stronger transcriptional activity did not involve different interactions. To quantify the strength of looping interactions, we calculated the intensity of the Hi-M map across an anchor to generate virtual interaction profiles (hereafter referred to as 4M plots). We observed that CRM_c

predominantly interacts with CRM_a and CRM_b with similar probabilities (Extended Data Fig. 2e vii). In contrast, we did not observe specific loops between CRMs and barcodes not containing early CRMs (e.g. ctrl barcode, Fig. 1c and Extended Data Fig. 2e iv). Interactions between CRMs and promoter regions (e.g. P1, P2 and P3) were present but displayed lower frequencies than interactions between CRMs (red arrow, Fig. 1f). Interactions between CRMs and the *doc1* promoter did not depend on the activation level of *doc1* (Extended Data Fig. 2f).

Next, we investigated whether all putative CRMs displayed chromatin loops. Interestingly, a CRM predicted from epigenetic profiling but not present in enhancer databases (e.g. CRM_a, see complete list of reported enhancers in Supplementary Table 1) displayed extensive interactions with reported enhancers (e.g. CRM_b, CRM_c, CRM_d) as well as with the promoters of *doc* genes (Extended Data Fig. 2e v). In contrast, a subset of barcodes harboring previously described enhancers, or displaying enhancer marks (see barcodes 2, 12, 13, 15 in Fig. 1c) failed to exhibit looping interactions with other CRMs (e.g. black circle in Fig. 1f). We observed similar results at the sna-TAD (Supplementary Note 1, Extended Data Fig. 2g,h). Thus, high-resolution Hi-M reveals unforeseen interactions between CRMs and other regulatory regions within the doc-TAD, and permits the quantification of the frequencies with which putative enhancers actively contact cognate target promoters in a specific tissue and developmental timing. Collectively, these data suggest that promoters interact with a panoply of enhancers that can be shared between different genes within a TAD.

Shared enhancers, promoter competition and CRM hubs

The existence of multiple pairwise interactions between CRMs within the doc-TAD and the naturally occurring overlapping expression patterns of *doc* genes (Extended Data Fig. 2a) suggests that multiple CRMs may compete or cooperate for gene activation in single cells[62]. To discriminate between these two hypotheses, we tested whether multi-way interactions are formed by excluding an anchor of interest and plotting the frequencies with which two barcodes interact together with this given anchor [31]. First, we selected promoters as anchors. We observed that 3-way interactions with multiple promoters were infrequent (green arrows, Fig. 1g i,ii,iii), consistent with our previous observations from Hi-M contact maps (Fig. 1f and Extended Data Fig. 2e). Instead, the three *doc* promoters preferentially looped to multiple CRMs in single nuclei (yellow arrows, Fig. 1g i,ii,iii). A control locus with no promoter marks failed to display specific looping interactions (Fig. 1g iv).

Genomic methods revealed the spatial clustering of multiple enhancers in cultured mammalian cells [30,31]. To test whether spatial clustering of multiple enhancers could be directly visualized in intact embryos, we mapped 3-way interactions using CRMs as anchors. Interestingly, we observed that $CRM_{a\text{-}d}$ displayed high frequencies of multi-way interactions (Fig. 1h, see examples labeled by yellow arrows). By analogy to previous studies [30,31], we termed these CRM interaction networks “CRM hubs”. CRM hubs can contain promoters (green arrows, Fig. 1h), but most often contained known or putative enhancers. Analysis of the sna-TAD reveals a similar scenario, where CRMs are involved in most 3-way interactions (Extended Data Fig. 2i iv, v and vi).

Finally, we used ShRec3D [63] to obtain an ensemble topological reconstruction from the Hi-M matrix. In this reconstruction CRMs clustered at the center of the TAD whereas promoter elements tended to be at the periphery (Fig. 1i). Similarly, we observed that CRMs within sna-TAD also tended to cluster together at the center of the TAD (Extended Data Fig. 2j). Collectively, 3-way and topological analyses suggest that multiple enhancers physically interact to form CRM hubs. Interestingly, CRM hubs can but do not tend to contain multiple promoters (Supplementary Note 2).

Networks of CRM contacts are indistinguishable between cells of different cell fates

Next, we examined whether chromatin structure in this locus depended on transcriptional status (repression/activation). For this, we used Hi-M in three populations of nuclei established along the dorso-ventral axis during nc14: mesoderm (M), neuroectoderm (NE) and dorsal ectoderm

(DE) [64]. To distinguish between these cell fates, we employed RNA-FISH labeling prior to Hi-M (with *sna* and *doc* probes directly labeling mesodermal and dorsal ectodermal cells, respectively, Fig. 2a). Nuclei were classified as: (a) dorsal ectoderm nuclei when an active *doc1* transcription hotspot could be visualized (Extended Data Fig. 2a); (b) mesoderm nuclei when located within the *sna* expression pattern (Fig. 2a); (c) neuroectoderm nuclei when located between the pattern of *sna* and the edge of the *doc1* pattern (Fig. 2a).

Unexpectedly, Hi-M interaction matrices for DE, NE and M displayed only minor differences (Fig. 2b-d), indicating that the same network of chromatin loops is present in nuclei that are actively transcribing and in nuclei where *doc* gene expression is silent. In addition, 4M profiles were almost identical in nuclei with different cell fates and activation status, independently of whether promoters or CRMs were used as anchors (Fig. 2e and Extended Data Fig. 3a,b). For instance, the *doc1* promoter (P1) showed identical interactions with the four CRMs (CRM_{a-d}) in the dorsal-ectoderm, the neuroectoderm and the mesoderm (Fig. 2e iii,vi). Likewise, CRM_a and CRM_c displayed patterns of interactions with other CRMs that were indistinguishable between tissues (Fig. 2e i,ii,iv,v). Finally, to detect whether CRM hubs existed in tissues where *doc* genes are repressed, we performed single-nucleus 3-way analyses. Indeed, comparison of 3-way interaction matrices of NE and M with those of DE revealed the persistence of CRM hubs in nuclei where transcription is repressed (Fig. 2f), suggesting that CRM hubs also exist in these cell types. We observed similar results in the *sna* locus (Supplementary Note 3, Extended Data Fig. 4a-d).

To search for a possible explanation of these results, we explored the transcription factor binding profiles of known activators and repressors in the *doc* locus [54,65,66] (Supplementary Note 4). CRM_{a-d} are bound by activators in the dorsal ectoderm, and tend to be occupied by transcriptional repressors in the mesoderm and neuroectoderm (Figure 2g-h). Thus, contacts between *doc* promoters and CRM hubs in the dorsal ectoderm may promote activation and those in the mesoderm/neuroectoderm may instead facilitate repression (Fig. 2h).

Cis-regulatory networks emerge before TADs and gene expression

Previous genome-wide and Hi-M studies have established that most *Drosophila* TADs emerge at nc14 during the major wave of zygotic gene activation (ZGA) [41,47,48]. To explore whether the doc-TAD also emerges at this nuclear cycle, we performed low-resolution Hi-M experiments in nc11-nc12 and nc14 embryos (Fig. 3a). We used nuclei density to unequivocally score developmental timing (Fig. 3b, insets). Hi-M contact maps revealed that the doc-TAD can be detected at nc14 but not at earlier stages (Fig. 3a), thus emergence of this TAD coincides with the onset of *doc* expression (Fig. 3c).

To determine whether specific looping interactions between CRMs appear before the emergence of TADs, we performed high-resolution Hi-M between nc11 and nc14. As our previous data showed that Hi-M maps are similar in different presumptive tissues (Fig. 2), we built Hi-M maps for the different nuclear cycles using all detectable nuclei independently of their location in the embryo. Surprisingly, chromatin loops between CRMs were observed very early in development (nc11) and remained almost unaffected at least until nc14 (Fig. 3b). For example, loops between CRM_c and CRM_a, CRM_b and CRM_d were readily detected as early as nc11, and assumed their final contact frequencies at nc12 (Fig. 3d). Similar behaviors were observed when using other CRMs as anchors (Extended Data Fig. 5a-c).

These results are consistent with 3-way analysis, where we observed that 3-way interactions are almost indistinguishable from nc12 to nc14 (Fig. 3e and Extended Data Fig 5d-f). To gather further evidence for the formation of CRM hubs during early development we obtained ShRec3D structures for each nuclear cycle. Notably, these structures show that CRMs cluster at the center of the TAD as early as nc11, with clusters becoming tighter as development progresses (Fig. 3f). We reached similar conclusions when analyzing the sna-TAD, which also emerges at nc14[48] (Supplementary Note 5, Extended Data Fig. 6). All in all, these data indicate that pairwise looping

interactions between CRMs in doc- and sna-TADs are established from nc11 (or before) while 3-way interactions are progressively acquired during development. Importantly, both pairwise and multi-way looping interactions are formed before the emergence of TADs.

To investigate whether specific loops between *doc* promoters and CRMs displayed quantitative changes before the onset of *doc* gene expression (*doc1* is the first to be activated, followed by *doc3* and *doc2*, Fig. 3c), we plotted virtual 4M profiles with promoters as viewpoints. Notably, we observed that promoters contact CRMs as early as nc11, and that frequencies of interactions ceased to change after nc12 (see P1 in Fig. 3g, and P2-P3 in Extended Data Fig. 5a-c). 3-way interactions involving promoters could also be already detected at nc11, and became more frequent at later nuclear cycles (Fig. 3h and Extended Data Fig. 5d-f). Thus, our results indicate that loops involving promoters and one or several CRMs precede TAD formation and gene expression, and are equally frequent in pluripotent cells, which do not express *doc* genes.

Redundancy of CRM usage at the doc-TAD

To test whether enhancers were redundant in this locus, we searched the literature[37,67,68] and performed enhancer reporter assays to identify the activation pattern of CRMs located around the *doc* genes (Methods). Several enhancers displayed partially overlapping patterns of activation at this stage of development, particularly in the region around CRM_c (Fig. 4a i-ii). This finding is consistent with multiple enhancers being able to activate the transcription of *doc* genes at this developmental stage.

To further test this hypothesis, we deleted CRM_c, which is located mid-way between two co-regulated promoters (*doc1* and *doc2*) (Methods, Fig. 4b and Extended Data Fig. 7a). Notably, deletion of CRM_c did not lead to detectable changes in *doc1* or *doc2* expression or on their co-activation frequency (Extended Data Fig. 7b), but produced large changes in the organization of the doc-TAD (Fig. 4c,d). Interactions involving the barcode originally containing CRM_c were considerably diminished, consistent with the binding of factors to this region being responsible for the formation of long-range CRM and promoter contacts (Fig. 4c,d, yellow arrows). Remarkably, we still observed interactions between the *doc1* promoter and the other enhancers (CRM_a, CRM_b, and CRM_d), indicating that removal of CRM_c did not affect the ability of the *doc1* promoter to be frequently in proximity to the other CRMs within the doc-TAD (Fig. 4c,d, red arrows). Finally, 3-way interactions between P1 and other CRMs (CRM_d, CRM_a, CRM_b) as well as with P2 persisted despite deletion of CRM_c (Fig. 4e, red arrows). Thus, CRM-CRM and CRM-P interactions still occur in the absence of CRM_c.

Formation of CRM hubs requires the pioneer factor Zelda

Having shown that interactions between multiple CRMs do not depend on transcriptional state or developmental timing, we searched for factors that may be required for the formation of CRM hubs. The pioneer factor Zelda (Zld) has the unusual ability to overcome nucleosome barriers at specific regulatory elements, making them accessible for binding by other classical TFs prior to activation, as early as nc8-nc11 [55–57,60,69,70]. The doc-TAD is enriched in Zld binding, particularly CRM_{a-d} (Fig. 4f), and Zld was required to ensure proper expression of *doc* genes, as well as for ensuring Pol 2 binding and chromatin accessibility at the *doc* locus (Supplementary Note 6, Extended Data Fig. 7c-e). To explore whether Zld depletion led to changes in the doc-TAD structure, we performed Hi-M experiments on Zld maternally depleted embryos using RNA interference (RNAi)[69]. Given the widespread developmental defects exhibited by *Zld* RNAi embryos at stage 5[71], we restricted our analyses to early nc14 *Zld RNAi* embryos.

Depletion of Zld did not affect TAD borders at the *doc* locus (Extended Data Fig. 7f), or changed the overall compaction of the doc-TAD (Extended Data Figs. 1i and 7g). However, we observed large changes in networks of CRM-CRM and CRM-promoter interactions (Fig. 4g). Contacts between CRM_a and other CRMs or promoters were very similar in Zld depleted and RNAi white

control embryos (Fig. 4g,h, red arrows; Supplementary Note 7). Remarkably, the main differences in contacts occurring in Zld-bound genomic regions that lose accessibility upon Zld depletion (CRM_c, CRM_d, CRM_b, P1, P2, P3) (Fig. 4g, yellow arrows; Fig. 4h, middle and right panels). For instance, upon Zld depletion, CRM_c showed the largest drop in ATAC-seq signal amongst CRMs (Fig. 4f), and a dramatic drop in its interactions with other CRMs (Fig. 4g,h, yellow arrows). Finally, in Zld RNAi embryos formation of CRM hubs was also considerably impacted (Fig. 4i) and topological reconstructions showed a loss of CRM clustering (Fig. 4j). Altogether, these results suggest a model whereby the pioneering activity of Zld plays a key role in the activation of *doc* genes and participates in the formation of CRM-CRM and CRM-promoter loops during early embryogenesis, possibly through its ability to open chromatin at specific CRMs.

To shed further light on the role of Zld in the formation of preferential interaction networks, we selected 5,038 genomic regions displaying Zld binding and calculated their pair-wise, intra-arm interaction frequencies [72] using publicly available datasets [41]. In nc14 embryos, Zld-bound regions interacted more frequently with each other than with control regions (Fig. 4k). This bias increased with the level of Zld binding and was present for short- (<250 kb) and long-range (>250 kb) genomic distances (Fig. 4k and Extended Data Fig. 7h). Zld depletion led to a considerable decrease in interactions between Zld-bound genomic regions (Fig. 4l and Extended Data Fig. 7i), consistent with Zld depletion results in the doc-TAD. However, this decrease in interactions was not observed upon transcriptional inhibition (Figs. 4m and Extended Data Fig. 7j), in agreement with other analyses [41]. Importantly, interactions between Zld-bound genomic regions were already present in nc12-nc13 embryos (Fig. 4n,o), and were also specific to Zld-bound pre-midblastula-transition (MBT) enhancers (Supplementary Note 8 and Extended Data Fig. 7k). Overall, these results are consistent with Zld being needed for the formation of a subset of CRM-CRM interactions during early embryogenesis.

Discussion

In this study, we use a high-resolution, imaging-based, single-cell spatial genomics approach (Hi-M) to link chromosome topology and transcriptional regulation during early *Drosophila* development. This approach has notable advantages, such as the detection of multi-way interactions and transcriptional output with spatial resolution (Supplementary Note 9). We reveal extensive interaction networks within developmental TADs primarily involving CRMs. Critically, these networks arise thanks to the spatial clustering of multiple enhancers (CRM hubs) and are mostly invariant during cell fate specification and gene activation. Networks of pairwise CRM contacts and CRM hubs arise during early development, before the onset of gene expression and before the emergence of TADs, and require the pioneering activity of the transcription factor Zld.

One of the important results of this study is that physical proximity between multiple CRMs and promoters is observed with very similar frequencies in cells with distinct fates and appeared during early embryogenesis. These results are consistent with those obtained at later stages of *Drosophila* embryogenesis, showing that enhancers located at considerably larger distances (~100 kb) can also form binary loops that are present in cells from different tissues [18]. Similarly, E-P interactions at the mouse *Hoxd* locus were detected in tissues where target genes were not expressed [19]. Our results are further supported by a companion paper [73] that applied Hi-C and micro-C to study tissue-specific *Drosophila* chromosome organization at similar stages of development (Supplementary Note 10). From a developmental perspective, the formation of loops between promoters and distal regulatory elements in cells where genes need to be repressed can be seen as a 'dangerous liaison'. Indeed, once a loop is established, transcriptional activation could rapidly occur in cells where that specific promoter should be kept inactive.

This apparent dichotomy, however, can be rationalized in terms of the spatiotemporal patterning of the *cis*-regulatory logic of transcription factors during embryogenesis. For instance, in mesodermal cells, most *doc* CRMs are bound by the spatially localized transcriptional repressor Sna

[45], which acts as a silencer in the mesoderm. In this case, communication between promoters and distal CRMs may reinforce transcriptional repression. This interpretation is in agreement with the finding that many enhancers can act as silencers in alternate cell types during *Drosophila* development [74], however other silencing mechanisms may also be at play [75]. Thus, we hypothesize that the optimal mechanism to ensure rapid and efficient activation or repression during development may involve two steps: the rapid priming of key CRMs via ubiquitously maternally deposited pioneer factors (e.g. Zld), followed by regulation of transcriptional output by spatially and temporally localized transcriptional activators and repressors. In this model, three-dimensional (3D) chromatin architecture plays a double role as 3D contacts could serve to reinforce either activation or repression at a particular developmental stage while allowing for flexibility at later stages. For instance, a repressive CRM loop in a tissue at an early developmental stage may switch to a CRM loop with activation capacities at later stages by changing transcription factor occupancy. Future experiments testing whether CRM loops and hubs display more differences in active and repressed tissues at later stages of development will be important to test these hypotheses.

Previous studies suggested that invariant E-P loops may be pre-established and stable [14,18,76,77]. In agreement with these results, our data indicate that E-P loops can form early, well before the onset of gene expression. However, in all cases, we measured low frequencies of looping interactions between functional elements. These results are consistent with previous measurements of absolute contact frequencies within TADs and between E-P [50,78–80]. Thus, these results indicate that different sets of multi-way E-E and E-P contacts are present in different cells, and that these contacts may be highly dynamic.

Recent studies reported the existence of enhancer hubs: spatially localized clusters containing multiple enhancers [30,31,34] that may facilitate transcriptional activation by creating a local microenvironment whereby transcriptional resources are shared, akin to early models of 'transcription hubs' [81]. Formation of enhancer hubs may require interactions between components of the transcriptional machinery, which could contribute to, or result from, the assembly of phase-separated condensates [33,38,39,82–84]. In this model, enhancers need not directly touch their target promoters but merely come into close proximity (~300 nm) [26,85]. Overall, these findings and models are consistent with our observation that multiple endogenous CRMs within a TAD come together in space to form hubs in single, actively transcribing nuclei. We also observed the formation of similar hubs in inactive nuclei, suggesting that repressive elements may also form spatially localized clusters of transcriptional repressors to share resources and reinforce their silencing activities. CRM hubs are formed at early stages of development in pluripotent cells. Thus, we favor a model in which preferential CRM interaction networks are pre-formed at early stages and are subsequently specified (into activation or repression hubs) during nc14 or later.

In *Drosophila*, TADs emerge concomitantly with the major wave of ZGA [41,47,48]. Previous studies reported the existence of chromatin loops typically at considerably large genomic distances spanning two or more TADs [18,47] or concerning Polycomb-binding sites [47,86]. Here, we observed that chromatin loops between CRMs within *Drosophila* TADs are widespread, mimicking the common CTCF-mediated chromatin loops present within mammalian TADs [16,42]. In addition, we found that multiple CRMs can cluster together to form *cis*-regulatory hubs located within TADs, suggesting a mechanism to sequester enhancers in space to reduce the activation of genes in neighboring TADs. Importantly, formation of CRM hubs precedes the emergence of TADs, consistent with the finding in mammalian cells that subsets of E-P contacts arise rapidly after mitosis before TADs are reformed [87]. Thus, our results suggest that CRM hubs and TADs likely form by different mechanisms. All in all, we hypothesize that CRM hubs represent an additional functional level of genome organization, independent of TADs. This additional layer can also be regulated by priming of enhancers and promoters by paused polymerases [88–90] or pioneer factors [55,56], as well as by chromatin marks [91]. As interactions between Zld CRMs appear before TADs, it is unclear how specificity of CRM interactions may be regulated to favor intra-TAD contacts (Supplementary Note 11).

Interestingly, we observed that interactions between Zld-bound CRMs, as well as interactions between CRMs and cognate promoters are established very early in pluripotent nuclei, prior to cell fate commitments. These long-range interactions occur between related CRMs (within doc- and sna- TADs) as well as between unrelated but Zld-bound CRMs (Fig. 4k and Extended Data Fig. 7k), suggesting that a common link could be their regulation by broad factors such as Zld. Critically, preferential contacts involving Zld-bound CRMs were considerably attenuated upon depletion of Zld. We and others have recently shown that Zld forms nuclear hubs in early *Drosophila* embryos [36,37], and that Zld hubs are re-established by the end of mitosis, prior to transcriptional activation. Taken together, our results suggest that Zld fosters the formation of CRM hubs by rendering chromatin accessible during early development, as a first step of cell specification to ensure maximum plasticity. Future work involving the detection of a larger number of CRMs will be needed to elucidate the factors and mechanisms involved in spatial clustering of developmental CRMs into nuclear micro-environments.

Acknowledgements

We are grateful to N. Benabdallah, G. Cavalli, T. Forne, T. Robert, J. Bonnet and members of the Lagha and Nollmann laboratories for their critical reading of the manuscript. We thank A. Makrini and D. Cattoni for help with bioinformatic analysis. This project was funded by the European Union's Horizon 2020 Research and Innovation Program (Grant ID 724429) (M.N.). We acknowledge the Bettencourt-Schueller Foundation for their prize 'Coup d'élan pour la recherche Française', the France-BioImaging infrastructure supported by the French National Research Agency (grant ID ANR-10-INBS-04, ''Investments for the Future''), and the *Drosophila* facility (BioCampus Montpellier, CNRS, INSERM, Université de Montpellier, Montpellier, France). M.G. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project ID 431471305. M.L.'s laboratory is supported by an ERC Starting Grant (SyncDev) and CNRS. M.B. and O.M. are supported by an FRM PhD fellowship.

Author Contributions

Conception and design: A.M.C.G., M.L., M.N.; Acquisition of data: S.M.E., C.H., M.B., I.R.; Analysis: M.G., S.M.E., M.B., O.M., I.R., M.N.; Software: M.G., M.N., J.B.F., O.M.; Interpretation of data: S.M.E., M.G., O.M., I.R., M.B., M.N., M.L.; Writing: M.L. and M.N.; Reagents: J.-B.F., C.H., I.R.; Visualization: S.M.E., M.G., M.N.; Supervision: M.N. and M.L.; Fund acquisition: M.N. and M.L.

Declaration of Interests

The authors declare no competing interests.

References

1. Bickmore, W. A. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
2. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
3. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
4. Alberts, B. *et al.* Molecular Biology of the Cell. (2017) doi:10.1201/9781315735368.
5. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
6. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358–368 (2016).

7. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* **62**, 237–247 (2016).
8. Schwarzer, W. & Spitz, F. The architecture of gene expression: integrating dispersed cis-regulatory modules into coherent regulatory domains. *Curr. Opin. Genet. Dev.* **27**, 74–82 (2014).
9. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
10. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
11. Ji, X. *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**, 262–275 (2016).
12. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
13. Ron, G., Globerson, Y., Moran, D. & Kaplan, T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat. Commun.* **8**, 2237 (2017).
14. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
15. Schwarzer, W. & Spitz, F. The architecture of gene expression: integrating dispersed cis-regulatory modules into coherent regulatory domains. *Curr. Opin. Genet. Dev.* **27**, 74–82 (2014).
16. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **162**, 687–688 (2015).
17. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
18. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).

19. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
20. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
21. Bartman, C. R. *et al.* Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. *Mol. Cell* **73**, 519–532.e4 (2019).
22. Morgan, S. L. *et al.* Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993 (2017).
23. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
24. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
25. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, (2019).
26. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* **76**, 473–484.e7 (2019).
27. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol. Cell* **77**, 688–708 (2020).
28. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
29. Enhancer-promoter communication: hubs or loops? *Curr. Opin. Genet. Dev.* **67**, 5–9 (2021).
30. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
31. Oudelaar, A. M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
32. Oudelaar, A. M. *et al.* A revised model for promoter competition based on multi-way

chromatin interactions at the α-globin locus. *Nat. Commun.* **10**, 5412 (2019).

33. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539–553.e8 (2020).
34. Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife* **8**, (2019).
35. Baudement, M.-O. *et al.* High-salt-recovered sequences are associated with the active chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies. *Genome Res.* **28**, 1733–1746 (2018).
36. Mir, M. *et al.* Dynamic multifactor hubs interact transiently with sites of active transcription in embryos. *Elife* **7**, (2018).
37. Dufourt, J. *et al.* Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat. Commun.* **9**, 5194 (2018).
38. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, (2018).
39. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842–1855.e16 (2018).
40. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife* **6**, (2017).
41. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216–228.e19 (2017).
42. Vallot, A. & Tachibana, K. The emergence of genome architecture and zygotic genome activation. *Curr. Opin. Cell Biol.* **64**, 50–57 (2020).
43. Porcher, A. & Dostatni, N. The bicoid morphogen system. *Curr. Biol.* **20**, R249–54 (2010).
44. Bonn, S. & Furlong, E. E. M. cis-Regulatory networks during development: a view of Drosophila. *Curr. Opin. Genet. Dev.* **18**, 513–520 (2008).
45. Stathopoulos, A. & Levine, M. Localized repressors delineate the neurogenic ectoderm

in the early Drosophila embryo. *Dev. Biol.* **280**, 482–493 (2005).

46. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
47. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell* **71**, 73–88.e5 (2018).
48. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* (2019) doi:10.1016/j.molcel.2019.01.011.
49. Cardozo Gizzi, A. M. *et al.* Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15**, 840–876 (2020).
50. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
51. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
52. Reim, I. & Frasch, M. The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in Drosophila. *Development* **132**, 4911–4925 (2005).
53. Hamm, D. C. *et al.* A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster. *PLoS Genet.* **13**, e1007120 (2017).
54. Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M. & Zeitlinger, J. Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.* **17**, 196 (2016).
55. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266 (2011).
56. Nien, C.-Y. *et al.* Temporal coordination of gene networks by Zelda in the early

Drosophila embryo. *PLoS Genet.* **7**, e1002339 (2011).

57. Hannon, C. E., Blythe, S. A. & Wieschaus, E. F. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife* **6**, (2017).
58. Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *Elife* **3**, (2014).
59. Rivera, J., Keränen, S. V. E., Gallo, S. M. & Halfon, M. S. REDfly: the transcriptional regulatory element database for Drosophila. *Nucleic Acids Res.* **47**, D828–D834 (2019).
60. Blythe, S. A. & Wieschaus, E. F. Establishment and maintenance of heritable chromatin structure during early embryogenesis. *Elife* **5**, (2016).
61. Senecal, A. *et al.* Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.* **8**, 75–83 (2014).
62. Hao, N., Shearwin, K. E. & Dodd, I. B. Positive and Negative Control of Enhancer-Promoter Interactions by Other DNA Loops Generates Specificity and Tunability. *Cell Reports* vol. 26 2419–2433.e3 (2019).
63. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).
64. Stein, D. S. & Stevens, L. M. Maternal control of the Drosophila dorsal-ventral body axis. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 301–330 (2014).
65. Deignan, L. *et al.* Regulation of the BMP Signaling-Responsive Transcriptional Network in the Drosophila Embryo. *PLoS Genet.* **12**, e1006164 (2016).
66. Van Bortle, K., Peterson, A. J., Takenaka, N., O'Connor, M. B. & Corces, V. G. CTCF-dependent co-localization of canonical Smad signaling factors at architectural protein binding sites in D. melanogaster. *Cell Cycle* **14**, 2677–2687 (2015).
67. Kvon, E. Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
68. Weiss, A. *et al.* A conserved activation element in BMP signaling during Drosophila development. *Nat. Struct. Mol. Biol.* **17**, 69–76 (2010).

69. Sun, Y. *et al.* Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
70. Schulz, K. N. *et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* **25**, 1715–1726 (2015).
71. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400–403 (2008).
72. Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A.-M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
73. Ing-Simmons, E. et al. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat. Genet.* (2021) [NG-A55249R1; accepted].
74. Gisselbrecht, S. S. *et al.* Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Mol. Cell* **77**, 324–337.e8 (2020).
75. Chopra, V. S., Kong, N. & Levine, M. Transcriptional repression via antilooping in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9460–9464 (2012).
76. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).
77. Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of during limb development. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12390–12399 (2019).
78. Cattoni, D. I. *et al.* Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* **8**, 1753 (2017).
79. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502–1515.e10 (2019).
80. Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-

regulatory elements. *Science* **359**, 1050–1055 (2018).

81. Jackson, D. A., Iborra, F. J., Manders, E. M. & Cook, P. R. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 (1998).
82. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
83. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, (2018).
84. Lesne, A., Baudement, M.-O., Rebouissou, C. & Forné, T. Exploring Mammalian Genome within Phase-Separated Nuclear Bodies: Experimental Methods and Implications for Gene Expression. *Genes* **10**, (2019).
85. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of Transvection in Living Drosophila Embryos. *Mol. Cell* **70**, 287–296.e6 (2018).
86. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. U. S. A.* (2017) doi:10.1073/pnas.1701291114.
87. Zhang, H. *et al.* Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* **576**, 158–162 (2019).
88. Ghavi-Helm, Y. *et al.* Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
89. Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* **153**, 976–987 (2013).
90. Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. & Ashe, H. L. Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. *Genes Dev.* **27**, 1146–1158 (2013).
91. Negre, N. *et al.* A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527–531 (2011).

Figure Legends

Fig. 1. Hi-M reveals widespread *cis*-regulatory chromatin loops and hubs within TADs

a. Schematic of the networks of contacts between *cis*-regulatory modules (CRMs) and promoters (P) within TADs.

b. The *doc* locus (Chr3L:8.88..9.03Mb) in *Drosophila melanogaster*. Low-resolution Hi-M and Hi-C [47] contact probability maps are shown on top and bottom, respectively. Blue and red indicate low and high contact probabilities, respectively.

c. Tracks for chromatin accessibility (ATAC-seq), Zelda binding, transcriptional activity (RNA-seq), chromatin marks for active promoters (H3K4me3), and active enhancers (H3K4me1, H3K27ac) from nc14 embryos, and RedFly enhancers are shown. Barcodes used for high-resolution Him included: regions with enhancer marks (CRM_{a-d}), *doc* promoters (P1-P3), intervening regions with no mark (e.g. 'ctrl'), and regions documented as enhancers (grey). See Supplementary Table 1 for assignment of CRM_{b-d}.

d. Schematic diagram of the labeling strategy.

e. Schematic representation of a dorsally oriented *Drosophila* embryo. Segmentation of actively transcribing nuclei (magenta) is based on nascent RNA FISH labeling.

f. The high-resolution Hi-M contact probability map of doc-TAD in nuclei displaying *doc1* expression in nc14 embryos. Boxes with irregular sizes above barcodes represent the relative genomic lengths of each barcode. Arrows: strong looping interactions between CRMs (yellow), expected promoter-promoter interaction (green), CRM-promoter interaction (red). Single-allele example reconstructions of spatially clustered and open CRM conformations. Statistics (panels f-h): N = 3,195 (nuclei with *doc1* expression), n = 24 (number of embryos with *doc1* expression), N_T = 37,129 (total number of nuclei), n_T= 29 (total number of embryos).

g. Multi-way interactions between promoter regions. Anchoring barcodes are highlighted by a pictogram. A control barcode is depicted in panel iv. Prominent peaks (yellow) comprise one promoter and two CRMs but not multiple promoters (green). Scheme illustrates the spatial arrangement of CRMs and promoter regions when the anchor is placed at a promoter.

h. Multi-way interactions between CRMs. Yellow arrows highlight prominent peaks involving three CRMs. Scheme illustrates the spatial arrangement of CRMs and promoter regions when the anchor is placed at a CRM.

i. 3D topological reconstruction of the doc-TAD.

Fig. 2. CRM-CRM and CRM-P loop frequencies are similar between cell types

a. Scheme indicating the three presumptive tissues and their segmentation (panel ii) based on RNA-FISH labeling (panel i).

b. Contact probability maps for dorsal ectoderm (DE) (upper-right half) and neuroectoderm (NE) (lower-left half) (panel i). Panel ii: map of the natural log of the ratio between contact probabilities of DE and NE. Blue indicates a larger contact probability in DE, red in NE.

c. Similar to panel b, but for DE and mesoderm (M).

d. Similar to panel b, but for NE and M.

e. 4M profiles derived from Hi-M maps for a selected number of anchors. Anchors were placed at CRM_a (panels i, iv), CRM_c (panels ii, v), and P1 (panels iii,vi).

f. Comparison of 3-way contacts for the same tissues and anchors as in panel e. Number of examined nuclei and embryos as in panel b, c, d, respectively.

g. ChIP profiles of key transcriptional regulators in the doc-TAD. Datasets are from whole embryos at nc14.

h. Illustration of the double role of CRMs in the doc-TAD.

Fig. 3. CRM loops and hubs precede TAD formation and gene expression

a. Low resolution Hi-M contact probability map of an extended genomic region around the doc-TAD. N: number of nuclei. n: number of embryos.

b. Representative images of DAPI-stained nuclei for embryos in nuclear cycles nc11 to nc14 (upper panel). High resolution Hi-M contact probability maps of the doc-TAD for embryos in nc11, nc12, nc13 and nc14. The minimum and maximum values of the linear color scale are indicated for each matrix using blue and red boxes.

c. Expression profile of *doc1*, *doc2* and *doc3* during nuclear cycles 10-14. Nuclear cycle 14 was divided into four time-points according to the extent of cellularization (a: earliest; d: last).

d. Comparison of 4M profiles derived from Hi-M maps at different nuclear cycles. The position of the anchor (CRM_c) is indicated by a vertical purple line. Profiles for nc11, 12 and 13 (orange lines) are compared to nc14 (blue lines) in panels i to iii, respectively.

e. Comparison of 3-way contacts between nc14 and other nuclear cycles, using CRM_c as anchor. Upper-right half of the matrix always depicts nc14, the bottom-left half shows nc11 (panel i), nc12 (panel ii) and nc13 (panel iii). Number of examined nuclei and embryos as indicated for the respective nuclear cycle in panel c.

f. Topological reconstructions of the doc-TAD for nc11 to nc14. CRMs and promoter regions are indicated as cyan and magenta spheres, respectively.

g. Similar to panel c, anchor: *doc1* promoter (P1)

h. Similar to panel d, anchor: *doc1* promoter (P1).

Fig. 4. Formation of CRM loops and hubs in the doc-TAD requires the pioneer factor Zelda

a. (i) Active enhancers in the doc-TAD and expression patterns: red represents pattern of activation of enhancers in the anterior-posterior axis of the DE, black represents the second sub-pattern of activation observed for *doc* enhancers. (ii) Representative images, as detected by lacZ RNA *in situ* hybridization, for the activation patterns of DocF5, DocF5SacIIb and DocF5SacIIa. Scale bar: 50 µm.

b. Simplified scheme for the conditional deletion of CRM_c.

c. Hi-M contact map for the CRM_c deletion mutant. Number of nuclei (N) and embryos (n) are the same for panels c-e.

d. 4M profiles for wild-type (yw) and CRM_c deleted embryos. Yellow arrows indicate missing interactions between CRM_c and other CRMs or promoters. Red arrows represent interactions between the *doc1* promoter and other CRMs.

e. Multi-way interactions in CRM_c deleted embryos. Red arrows represent interactions between the *doc1* promoter and other CRMs/promoters.

f. ATAC-seq profiles of wildtype, *Zld* mutant, and $\log_2$ ratio between *zld*- and *wt* nc14 embryos. Barcodes are annotated below. Arrows: peaks showing a decrease in chromatin accessibility upon Zld depletion.

g. High-resolution Hi-M contact probability maps for *white-RNAi* (upper-right matrix) and *zld-RNAi* (bottom-left matrix). Yellow arrows indicate missing interactions between CRMs and other CRMs/promoters. Red arrows represent the conserved interactions between CMRs. Number of nuclei (N) and embryos (n) are the same for panels g-i.

h. 4M profiles for *white-RNAi* and *zld-RNAi* embryos.

i. Multi-way interactions in *white-RNAi* (upper-right map) and *zld-RNAi* (bottom-left map) embryos.

j. Topological reconstructions of doc-TAD for *wt* and *zld-RNAi* embryos (left, center). The arrow indicates a separation of CRMs in the absence of Zld. Single-allele structure from a *zld-RNAi* embryo (right).

k-o. Log_2(observed/expected) average contact frequencies between Zelda-bound regions at short-range distances (≤250 kb) ranked by increasing Zelda enrichment in nc14 (panel k), nc14 *zld*-RNAi (panel l), nc14 alpha-amanitin-treated (panel m), nc12 (panel n), and nc13 (panel o) embryos.

Online Methods

Drosophila stocks and embryo collection

Fly stocks were maintained at room temperature (RT) with natural light/dark cycle and raised in standard cornmeal yeast medium. The *yw* or UASp-shRNA-w (BL#35573) stocks were used as a control. Flippase stock (BL#26902) and CRE stock (BL#851) were used for the generation of the CRMc deletion strain. Zelda-depleted embryos were obtained from females from the cross between *nos-Gal4*:VP16 (BL#4937) and *UASp-shRNA-zld* [69], a method with a Zld depletion efficiency of ~90% [37]. After a pre-laying step, flies were allowed to lay eggs for 1.5 h on new yeasted 0.4% acetic acid plates. Embryos were then incubated at 25 °C until they reached the desired developmental stage. Embryos were collected and fixed as previously described [49]. Briefly, embryos were dechorionated with 2.6 % bleach, rinsed and fixed with a 1:1 mixture of 4% methanol-free formaldehyde in PBS and Heptane. Embryos were stored in methanol at -20 °C until further use.

Generation of the CRM_c deletion and reporter lines

A CRISPR/Cas9-based strategy was employed to specifically delete CRM_c in a conditional manner. Two FRT sites were inserted flanking a ~860-bp region (3L:9021947-9022805) surrounding the most prominent Zld peaks in CRM_c (impacting barcodes 13-14). Recombination template was cloned into pHD-DsRed plasmid (Addgene #51434). First, a 5' homology arm (PCR-amplified from genomic DNA) was inserted into a vector previously digested with XmaI/NheI. Then, PCR-amplified 3' homology arm was inserted after digestion with SpeI/AscI. Finally, PCR-amplified CRM_c flanked by FRT sequences was cloned after digestion by NotI. Guide RNAs were cloned into pCFD3-dU6:3gRNA (Addgene #49410) digested by BbsI using annealed oligonucleotides. Recombination template and guide RNAs were injected by BestGene Inc. After obtaining the CRISPR-edited stocks, males were crossed with *CRE/CRE; D*/TM3,Sb* virgin females to remove dsRed marker by the action of a Cre recombinase. Then, *dsRED-/-* males were crossed with *Dp/TM3,Sb* virgins females. Males from this cross were then crossed with *hs-FLP/hs-FLP;Dr/TM3,Sb* virgins females. Larvae from these crosses where heat shocked at 37 °C for 30 min in a water bath in order for the flippase to be expressed and delete the CRM_c. Adult males were then PCR genotyped. Oligonucleotide sequences used for cloning and genotyping are listed in Supplementary Table 2. The deletion removes 287 bp of barcode 13 (5 oligonucleotides out of 50) and 562 bp (10 oligonucleotides out of 50) of barcode 14.

DocFX-lacZ reporter lines were generated by P transgenesis of pH-Pelican vectors with CRM fragments in between KpnI and NotI sites analogous to the procedure described in Kahn et al.[94]. Subfragments F5SacIIa and F5SacIIb were generated by removing KpnI-SacII or SacII-NotI fragments, respectively, with subsequent blunt end religation.

Hi-M libraries

Oligopaint libraries, consisting of unique ~35/41-mer sequences with genome homology, were obtained from the Oligopaint public database (http://genetics.med.harvard.edu/oligopaints). We selected 20 barcodes in the *doc* locus (3L: 8882600..9039000 *Drosophila* release 5 reference genome in all cases) for the low-resolution Hi-M library, 17 barcodes encompassing the doc-TAD (3L: 8974562..9038920) for the high-resolution Hi-M library, and 65 barcodes (2L:15244500..15630000) for the high-resolution *sna* locus library. For each barcode, we used 45-50 probes, covering ~3 kb. An additional fiducial barcode located at least ~1 Mb away was used for drift correction (see below). The coordinates of the targeted genomic regions are listed in Supplementary Table 3.

Each oligonucleotide in the pool consisted of 5 regions: i- a 21-mer forward priming region, ii- a 32-mer (low-res library) or two 20-mer separated by an AT sequence (high-res libraries) readout region unique for each barcode, iii- a 35/41-mer genome homology region, iv- a 32-mer (low-res library) or 20-mer (high-res libraries) readout region and v- a 21-mer reverse priming region. The designed template oligonucleotide pools were ordered from CustomArray. The procedure to amplify oligonucleotide pools to obtain the primary libraries was as previously described [49]. It involved a 5-step procedure consisting of i- limited-cycle PCR, ii- amplification via T7 *in vitro* transcription, iii- reverse transcription and iv- alkaline hydrolysis and purification. The sequence of the primers used for amplification of the libraries are listed in Supplementary Table 4.

For the low-resolution library, we employed 20 unique Alexa647-labeled sequence oligonucleotides (imaging oligonucleotides), complementary to the readout region present in the primary oligonucleotide. The fluorophore was attached via a disulfide linkage cleavable by the mild reducing agent Tris(2-carboxyethyl)phosphine (TCEP), using a previously described strategy [49]. Alternatively, for the high-resolution libraries, we used "adapter" oligonucleotides, consisting in a 20-mer region complementary to the readout sequence able to recognize the barcode being targeted, a 10-mer spacer sequence and a 32-mer region able to bind to a unique Alexa647-labeled oligonucleotide (containing a disulfide linkage). In this approach, a single fluorescent oligonucleotide is required [50]. For fiducial barcodes, a non-cleavable Rhodamine-labeled oligonucleotide was used. The sequences of the imaging and adapter oligonucleotides are listed in Supplementary Table 5. PCR and reverse transcription primers used in probe synthesis, as well as adapter oligonucleotides and fluorescently labeled oligonucleotides, were purchased from Integrated DNA Technology (IDT). The whole set of Oligopaints used can be found in Supplementary Table 6.

RNA-FISH probes

RNA probes were obtained by *in vitro* transcription from a vector containing the sequences targeting *sna* (previously described in [48]), *doc1*, *doc2* or *doc3* genes in the presence of digoxigenin (DIG) or biotin (BIO) haptenes. Vector was linearized before the *in vitro* transcription with a specific restriction enzyme. RNA probes produced in this manner were then treated with carbonate buffer at 65 °C for 5 min (*sna* probe) or for 2 min (*doc1,doc2,doc3* probes). The information on each probe, including the primers used to clone the target sequences by amplification of genomic DNA, are listed in Supplementary Table 7.

RNA Fluorescent *In situ* Hybridization

In situ hybridization was as described previously [49], with modifications to allow for the detection of two different species of RNA. The reader is invited to read our detailed protocol in the aforementioned reference. Briefly, fixed embryos were passed through 1:1 mixture of methanol:ethanol and then pure ethanol. Embryos were then post-fixed with 5% formaldehyde in PBT (PBT = 0.1% Tween-20 PBS) for 25 min. Then, embryos were incubated 4 times with PBT during

15 min and permeabilized 1 h with 0.3% Triton in PBS. Embryos were rinsed with PBT and incubated for 2 h with RHS at 55 °C (RHS = 50% formamide, 2× SSC, 0.1% Tween-20, 0.05 mg/ml heparin, 0.1 mg/ml salmon sperm). In the meanwhile, RNA probes were heated at 85 °C for 2 min, transferred to ice for 2 min and then incubated with the embryos in RHS for 16-20 h at 55 °C for RNA hybridization. The next day, embryos were washed 4 times with RHS at 55 °C and 3 times with PBT at RT. Then, a saturation step was performed with blocking solution (blocking reagent Sigma #11096176001, 100 mM Maleic acid, 150 mM NaCl, pH = 7.5) for 45 min.

Then the protocol depends on whether embryos were used for Hi-M (*sna*/*doc1* double labeling) or to reveal *doc1, doc2* and *doc3* expression patterns (Extended Data Fig. 1j). To reveal the expression patterns of *doc* genes, the combination of *doc1*-DIG/d*oc2*-BIO or *doc2*-DIG/*doc3*-BIO was used. After the saturation step, embryos were incubated with primary antibodies at 1:375 dilution (sheep anti-DIG, Roche cat #11333089001 and mouse anti-Biotin, Life technologies cat #03–3700) overnight at 4°C. The next day embryos were washed 6 times in PBT for 10 min. Embryos were incubated 1 h in blocking solution, then 2 h with secondary antibodies at 1:500 dilution (anti-mouse Alexa488-conjugated, Life technologies cat #A21202 and anti-sheep Alexa555-conjugated, Life technologies cat #A21436) and washed 6 times in PBT. Finally, embryos were incubated 10 min with a 0.5 mg/ml DAPI solution, washed with PBT and mounted in ProLong™ Diamond Antifade.

For Hi-M, both *sna* and *doc1* probes were DIG-labeled. By taking advantage of the differential spatial expression pattern, we labeled both RNAs simultaneously by the combination of both probes during incubation and the use of a single anti-DIG antibody and a tyramide signal amplification (TSA) reaction. After RNA hybridization and the saturation step, the activity of endogenous peroxidases was eliminated by incubating with 1% H_2O_2 in PBT for 30 min. After rinsing with PBT, embryos were incubated overnight at 4 °C with sheep anti-DIG conjugated with POD (Sigma-Aldrich cat #11207733910) with 1:500 working dilution in PBT. The next day, embryos were washed with PBT and incubated for 30 min with tyramide-coupled Alexa 488. Next, H_2O_2 was added to a final concentration of 0.012% during another 30 min. Embryos were washed with PBT and stored at 4 °C until further use.

Hybridization of Hi-M primary library

Hybridization followed a previously described protocol [49]. Briefly, embryos were RNase-treated for 2 h, permeabilized 1 h with 0.5% Triton in PBS and rinsed with sequential dilutions of Triton/pHM buffer to 100% pHM (pHM = 2× SSC, NaH_2PO_4 0.1 M pH = 7, 0.1% Tween-20, 50% formamide (v/v)). Embryos in pHM were preheated at 80 °C, the supernatant was aspirated and 30 µl of FHB (FHB = 50% formamide, 10% dextran sulfate, 2× SSC, salmon sperm DNA 0.5 mg/ml) containing 225 pmol of the primary library was pipetted directly onto the embryos. Mineral oil was added on top and the tube was incubated overnight at 37 °C. The next day, oil was carefully removed and embryos were washed two times during 20 min at 37 °C with 50% formamide, 2× SSC, 0.3% CHAPS. Next, embryos were sequentially washed for 20 min at 37 °C with serial dilutions of formamide/PBT to 100% PBT. Embryos were rinsed with PBT and stored in PBT at 4 °C until the imaging step.

Imaging system

All experiments were performed on a home-made wide-field epifluorescence microscope built on a RAMM modular microscope system (Applied Scientific Instrumentation) coupled to a microfluidic device as described previously [48,49]. Samples were imaged using a 60× Plan-Achromat water-immersion objective (NA = 1.2, Nikon, Japan). The objective lens was mounted on a closed-loop piezoelectric stage (Nano-F100, Mad City Labs Inc., USA). Illumination was provided by 4 lasers (OBIS-405/488/640 nm and Sapphire-LP-561 nm, Coherent, USA). Images were acquired using a sCMOS camera (ORCA Flash 4.0V3, Hamamatsu, Japan), with a final pixel size calibrated to 106 nm. A

custom-built autofocus system was used to correct for axial drift in real-time and maintain the sample in focus as previously described [48].

A fluidic system was used for automated sequential hybridizations, by computer-controlling a combination of three eight-way valves (HVXM 8-5, Hamilton) and a negative pressure pump (MFCS-EZ, Fluigent) to deliver buffers and secondary readout probes onto a FCS2 flow chamber (Bioptechs). Software-controlled microscope components, including camera, stages, lasers, pump, and valves were run using a custom-made software package developed in LabView 2015 (National Instrument).

Acquisition of Hi-M datasets

Embryos were attached to a poly-L-lysine coated coverslip and mounted into the FCS2 flow chamber. Fiducial readout probe (25 nM Rhodamine-labeled probe, 2× SSC, 40% v/v formamide) was flowed onto the sample and hybridized for 15 min, washed for 10 min with readout washing buffer (2× SSC, 40% v/v formamide) and for 5 min with 2× SSC before injecting 0.5 mg/ml DAPI in PBS to stain nuclei. The imaging buffer (1x PBS, 5% w/v glucose, 0.5 mg/ml glucose oxidase and 0.05 mg/ml catalase) was injected. Subsequently, 10-15 embryos were selected according to developmental stage and orientation and segmented into a mosaic of multiple fields of view (FOV of 200 × 200 µm). After bright field image recording, z-stacks were taken with 405, 488 and 561 nm laser illuminations. The z-stacks had a step size of 250 nm with a total range of 15 µm.

Next, the sample was sequentially hybridized with different secondary readout probes, imaged in the Rhodamine and the Alexa-647 channels, and photobleached. For each round of secondary hybridization, the sample was treated with secondary hybridization buffer (25-50 nM imaging oligonucleotide, 2× SSC, 40% v/v formamide, that also included 50 nM of adapter oligonucleotide in the case of the high-res libraries, see *Hi-M libraries*,) for 15 min, then washed with readout washing buffer and with 2× SSC before injecting imaging buffer. After imaging, the fluorescence of the readout probes was extinguished using a chemical bleaching buffer (2× SCC, 50 mM TCEP hydrochloride) for 10 min and then the sample was washed with 2× SSC for 5 min before a new hybridization cycle started. All buffers were freshly prepared and filtered for each experiment. The imaging buffer used for a single experiment was stored under a layer of mineral oil and renewed every 12-15 h. Further details can be found on our previously published protocol [49].

Image processing

Our home-made Hi-M microscope produced z-stacks in DCIMG format, which were converted to TIFF using proprietary software from Hamamatsu. TIFF images were then deconvolved using Huygens Professional version 20.04 (Scientific Volume Imaging, the Netherlands, https://svi.nl/). Further analysis steps were performed using a homemade analysis software that implemented the steps described previously [49]. Briefly, images were first z-projected using either sum (DAPI channel) or maximum intensity projection (barcodes, fiducials). Image-based cross-correlation was used to align the fiducial channels. These corrections were then used to align DAPI and barcode images. Next, the positions of the XY centers of barcodes were detected with sub-pixel resolution using local maximum fitting functions from the ASTROPY package [95]. Nuclei were segmented from projected DAPI images by adaptive local thresholding and watershed filtering [49]. RNA images were segmented by manually drawing polygons over the nuclei displaying a pattern of active transcription. Barcodes and RNA status were then attributed to each single nuclei by using the XY coordinates of the barcodes, the projected DAPI masks of nuclei, and the transcriptional status from manual masking. Finally, pairwise distance matrices were calculated for each single nucleus. From the list of pairwise distances obtained from any two barcodes, we calculated the contact probability as the number of nuclei in which pairwise distances were within 250 nm normalized by the number of nuclei containing both barcodes. This definition was similar to that used in other studies [78,79], and avoided biases due to uneven barcode detections. Contact frequencies obtained using this pipeline and those using

previous pipelines [49] produced highly correlated results. Image processing was carried out from Linux terminals connected to a server running Linux PopOS 19.10, with 4 GeForce GTX 1080Ti GPU cards (SCAN computers, UK). Assessment of Hi-M dataset size was done using a bootstrapping approach (Extended Data Fig. 1h).

4M profiles and multi-way interactions

4M profiles were obtained by slicing the corresponding Hi-M contact map across a given anchor. Multi-way interactions were obtained by selecting an anchoring barcode and calculating the single-nucleus pairwise distances to all possible pairs of barcodes. If both barcode-anchor distances for a given barcode pair in a single nucleus are below the contact threshold (250 nm), this nucleus is considered to have a 3-way interaction for this anchor and barcode pair. The 3-way contact frequency is then obtained by dividing the number of nuclei that show a 3-way interaction by the number of nuclei where the three barcodes involved in the 3-way interaction have been detected. The calculation of 3-way interactions does not require the detection of contiguous barcodes, therefore the calculation of 3-way frequencies is not restricted to nuclei displaying all the barcodes.

ShRec3D Structures

Three-dimensional topological representations were obtained from Hi-M pairwise distance maps using our own Python implementation of the approaches described by Lesne *et al.* and Morlot *et al.* for ShRec3D [63,96]. Starting from the single-cell pairwise distance matrix, an ensemble pairwise distance matrix was calculated using the first maximum of the kernel density estimation. These pairwise distances were converted into 3D coordinates for each barcode using nonclassical metric multidimensional scaling. When necessary, structures were mirrored and a ball-and-stick representation was rendered with PyMOL (The PyMOL Molecular Graphics System, Version 2.3 Schrödinger, LLC.).

Genome-wide analysis of Zld-mediated interactions

First, we extracted lists of Zld peaks genome-wide. For Zld peaks used in Figure 4k-o, Extended Data Figure 7i-k: Datasets from Harrison et al. [55] (GSM763062) were used to extract the autosomal regions bound by Zelda at 3 hpf using their corresponding ChIP-Seq intensity [55]. For Zld peaks used in Extended Data Figure 7l: a list of putative enhancers of pre-MBT genes (N = 62) was obtained by selecting the Zld peaks nearest to the TSS of pre-MBT genes [97] (Supplementary Table 8). BED coordinates were remapped from dm3 (BDGP R5) to dm6 (BDGP R6) using FlyBase's sequence coordinates converter (FB2020_05, released Oct 14, 2020).

Second, we characterized interactions between Zld-bound regions using the 5-kb resolution Hi-C dataset from Hug. *et al*. [41]. Peaks were sorted and classified in different categories based on protein occupancy. If multiple Zelda peaks were contained within the same 5-kb bin, only the one with the highest intensity was considered. After filtering, we analyzed 5,038 bins occupied by Zelda in the different autosomal chromosomes. For each biological condition, intra-arm chromosomal contacts were distance normalized by computing the $\log_2$ [observed/expected]. The average interaction frequencies at long- (>250 kb) or short-ranges (≤ 250 kb) were then ranked in 5 groups by increasing Zelda ChIP signal. From low to high Zelda peak intensity each group contains respectively 3,124, 706, 480, 248, and 480 peaks.

Data availability

Oligopaint public database (http://genetics.med.harvard.) was used to select oligopaints. Publicly available datasets used in this study (GSE86966, GSE25180,E-MTAB-4918, GSM763062, GSE58935, GSE16245, GSE68983, GSE68654, E-MTAB-1673, GSE62904, GSE65441) are detailed in Supplementary Table 9. Uncropped version of agarose gel in Extended Data Figure 7a is provided in

Source Data Extended Data Figure 7. Data for matrices in Figures 1-4 and in Extended Data Figures are publicly available at https://github.com/NollmannLab/Espinola-Goetz-2021.

Code availability

Code used in this manuscript is available at https://github.com/NollmannLab/Espinola-Goetz-2021.

Methods-only References

92. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
93. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* (2016) doi:10.1038/nature16496.
94. Kahn, T. G. *et al.* Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res.* **44**, 10132–10149 (2016).
95. The Astropy Collaboration *et al.* The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJS* **156**, 123 (2018).
96. Morlot, J.-B., Mozziconacci, J. & Lesne, A. Network concepts for analyzing 3D genome structure from chromosomal contact maps. *EPJ Nonlinear Biomedical Physics* vol. 4 (2016).
97. Chen, K. *et al.* A global change in RNA polymerase II pausing during the Drosophila midblastula transition. *Elife* **2**, e00861 (2013).

Figure 1

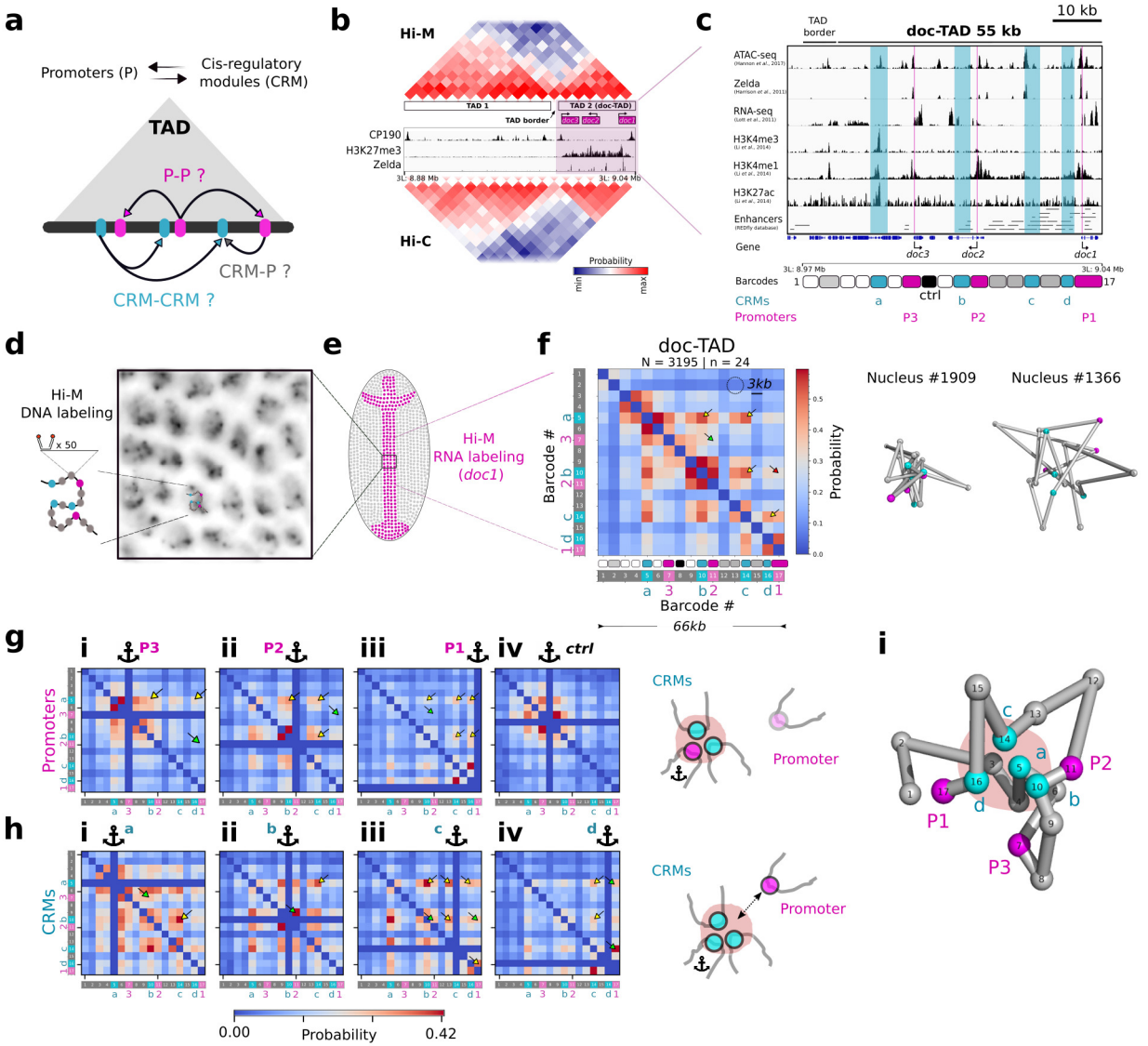

a
Promoters (P)
Cis-regulatory modules (CRM)
TAD
P-P ?
CRM-P ?
CRM-CRM ?
b
Hi-M
TAD 1
TAD 2 (doc-TAD)
TAD border
doc3
doc2
doc1
CP190
H3K27me3
Zelda
3L: 8.88 Mb
3L: 9.04 Mb
Hi-C
Probability
min
max
c
TAD border
doc-TAD 55 kb
10 kb
ATAC-seq
Zelda
RNA-seq
H3K4me3
H3K4me1
H3K27ac
Enhancers
Gene
doc3
doc2
doc1
3L: 8.97 Mb
3L: 9.04 Mb
Barcodes
1
17
ctrl
CRMs
a
b
c
d
Promoters
P3
P2
P1
d
Hi-M
DNA labeling
x 50
e
Hi-M
RNA labeling
(doc1)
f
doc-TAD
N = 3195 | n = 24
3kb
Barcode #
Probability
Nucleus #1909
Nucleus #1366
66kb
g
i
P3
ii
P2
iii
P1
iv
ctrl
Promoters
CRMs
Promoter
h
i
a
ii
b
iii
c
iv
d
CRMs
0.00
Probability
0.42
i
c
a
P2
b
d
P1
P3

Figure 2

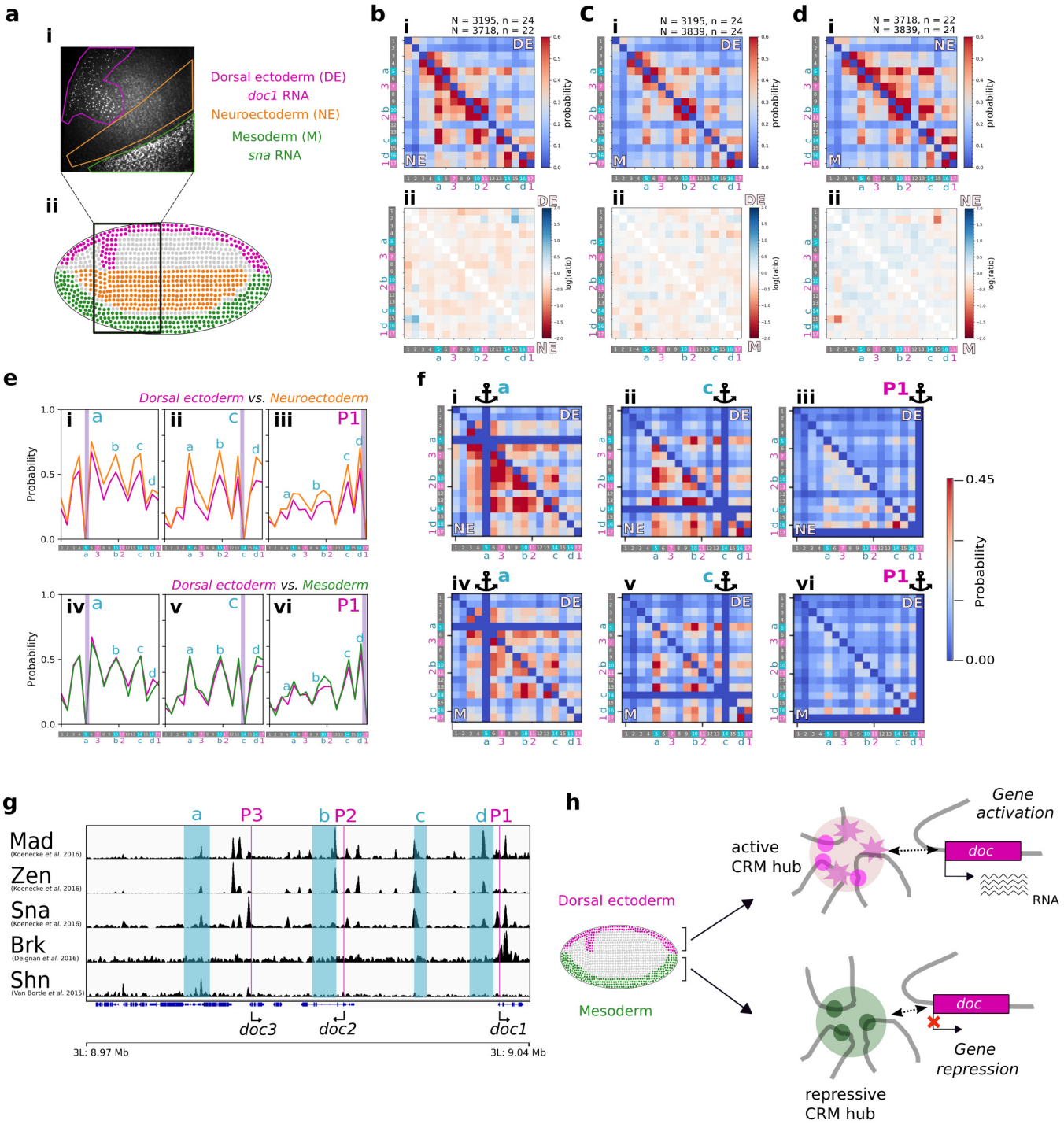

a
i
ii
Dorsal ectoderm (DE)
doc1 RNA
Neuroectoderm (NE)
Mesoderm (M)
sna RNA
b
c
d
N = 3195, n = 24
N = 3718, n = 22
N = 3839, n = 24
N = 3718, n = 22
probability
log(ratio)
DE
NE
M
e
Dorsal ectoderm vs. Neuroectoderm
Dorsal ectoderm vs. Mesoderm
Probability
f
P1
0.45
0.00
g
Mad
Zen
Sna
Brk
Shn
(Koenecke et al. 2016)
(Deignan et al. 2016)
(Van Bortle et al. 2015)
P3
P2
doc3
doc2
doc1
3L: 8.97 Mb
3L: 9.04 Mb
h
Dorsal ectoderm
Mesoderm
active CRM hub
repressive CRM hub
Gene activation
Gene repression
doc
RNA

Figure 3

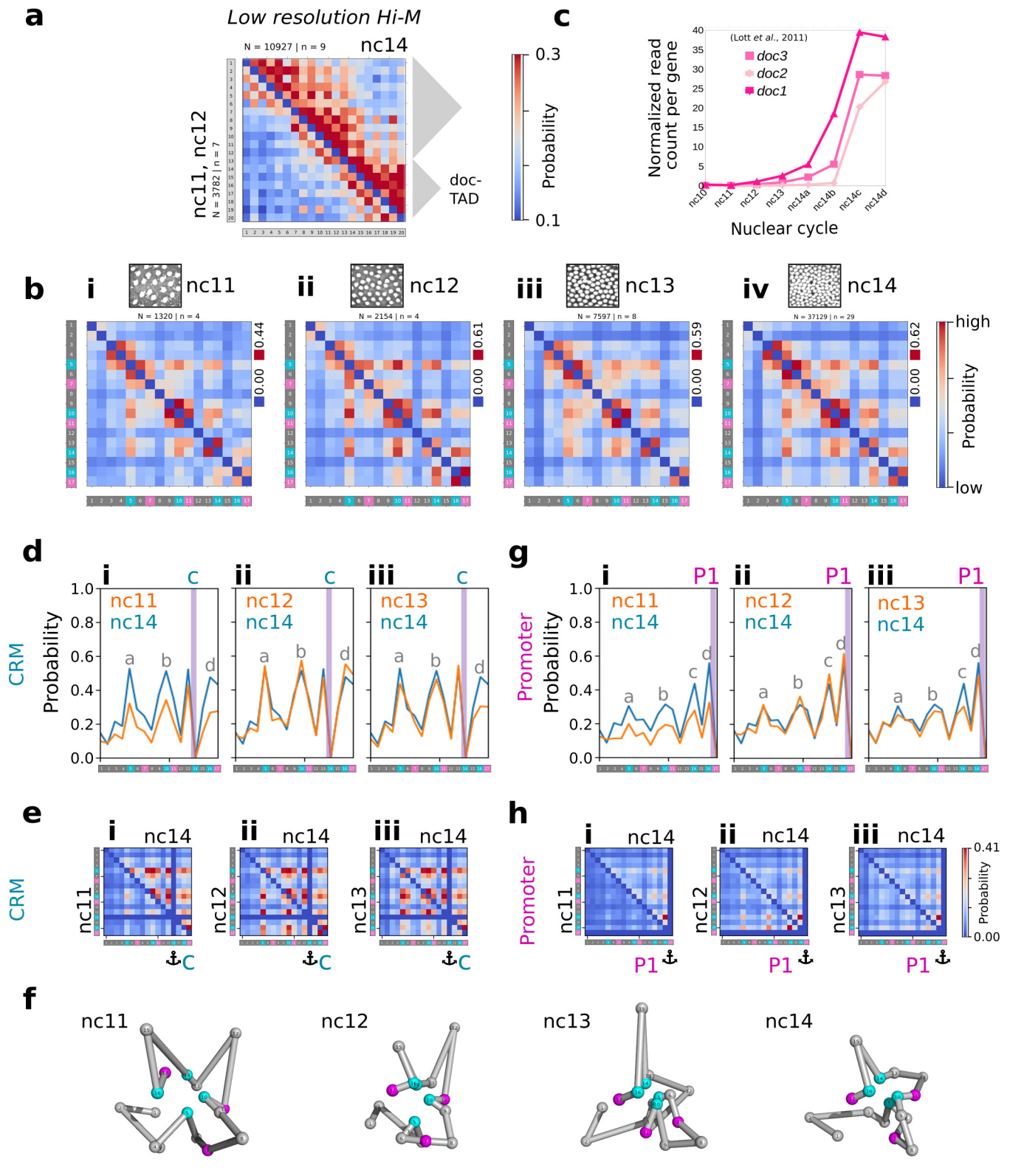
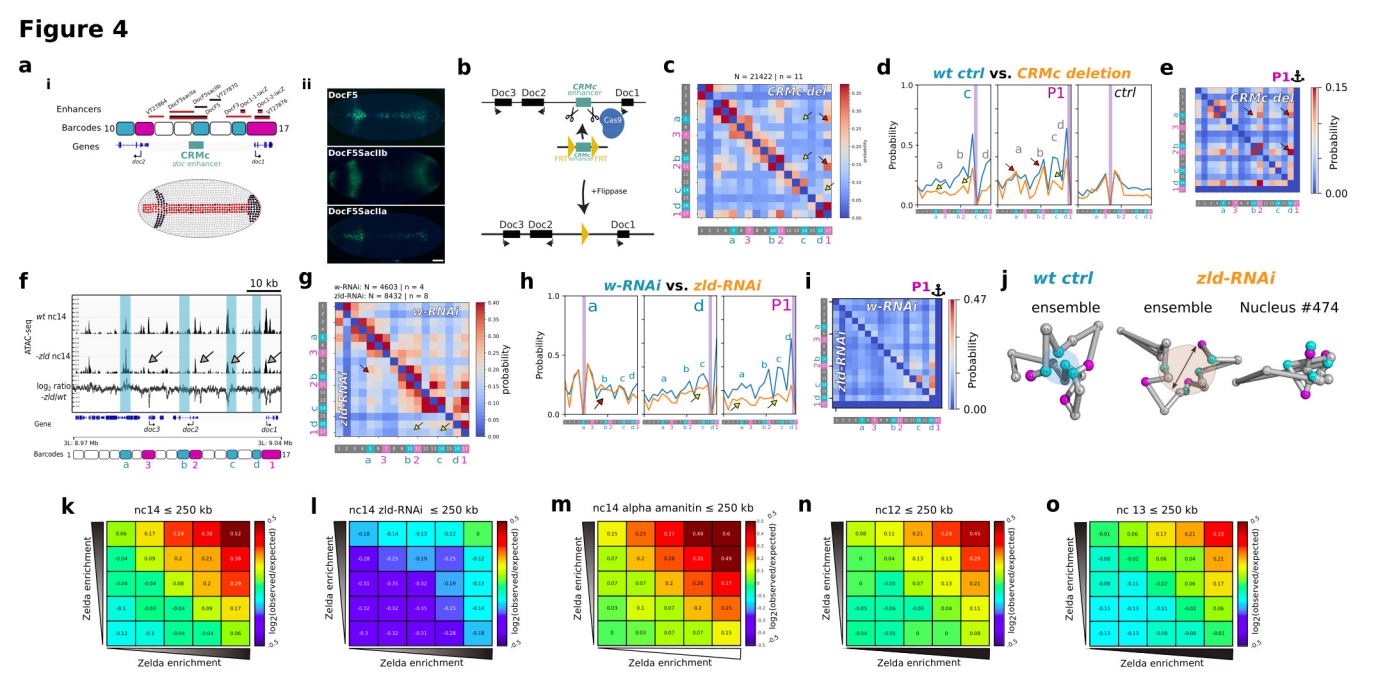
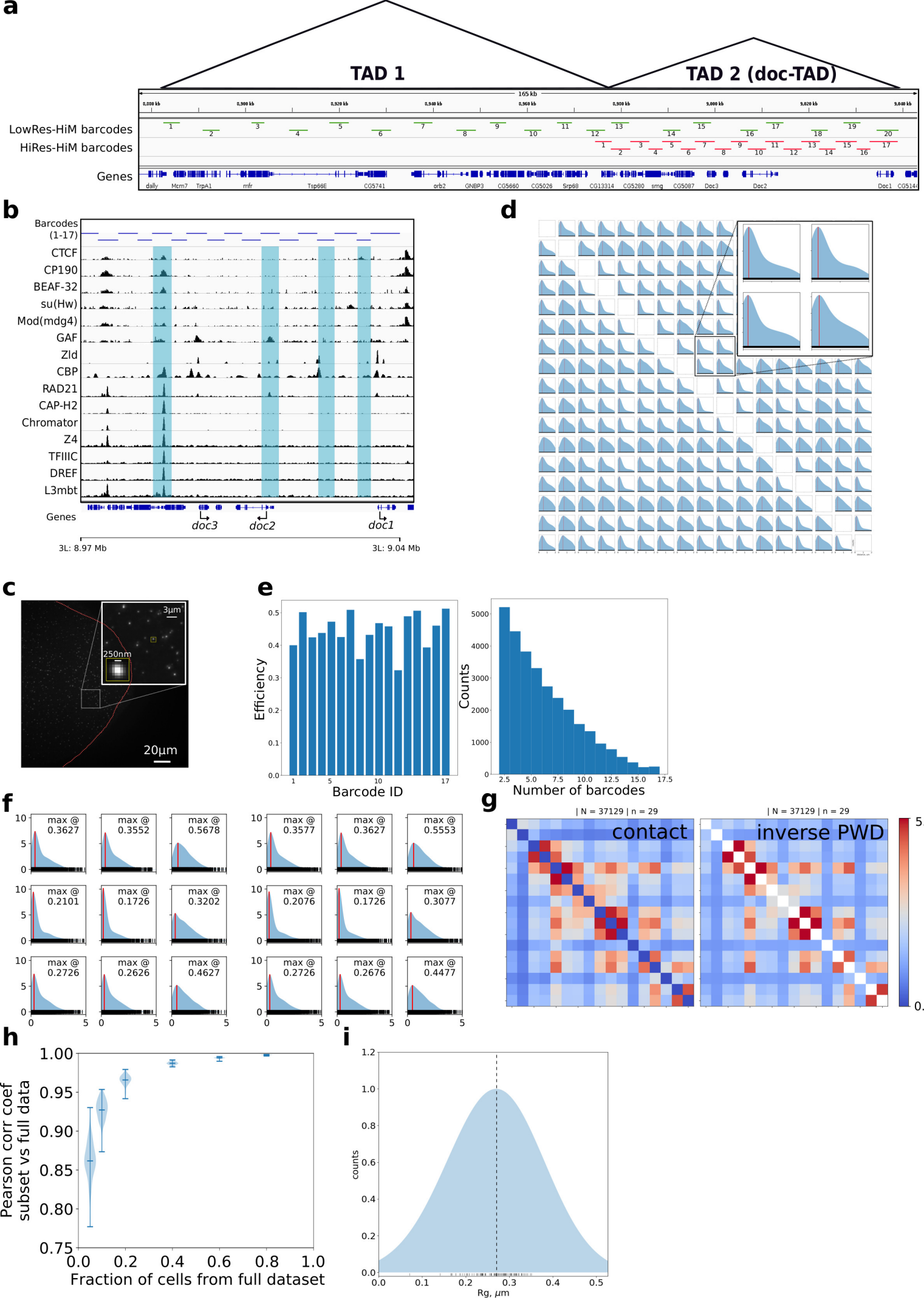
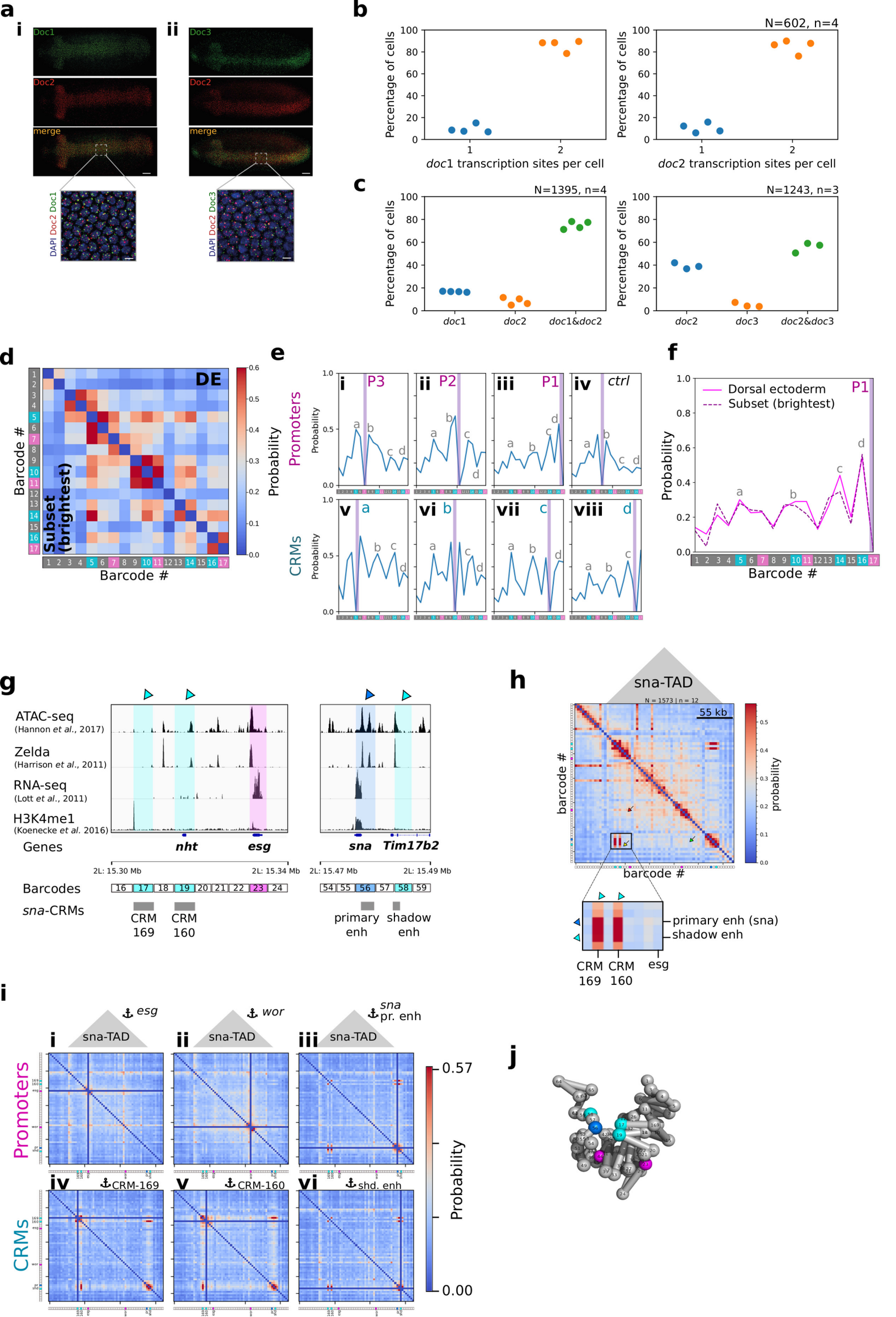
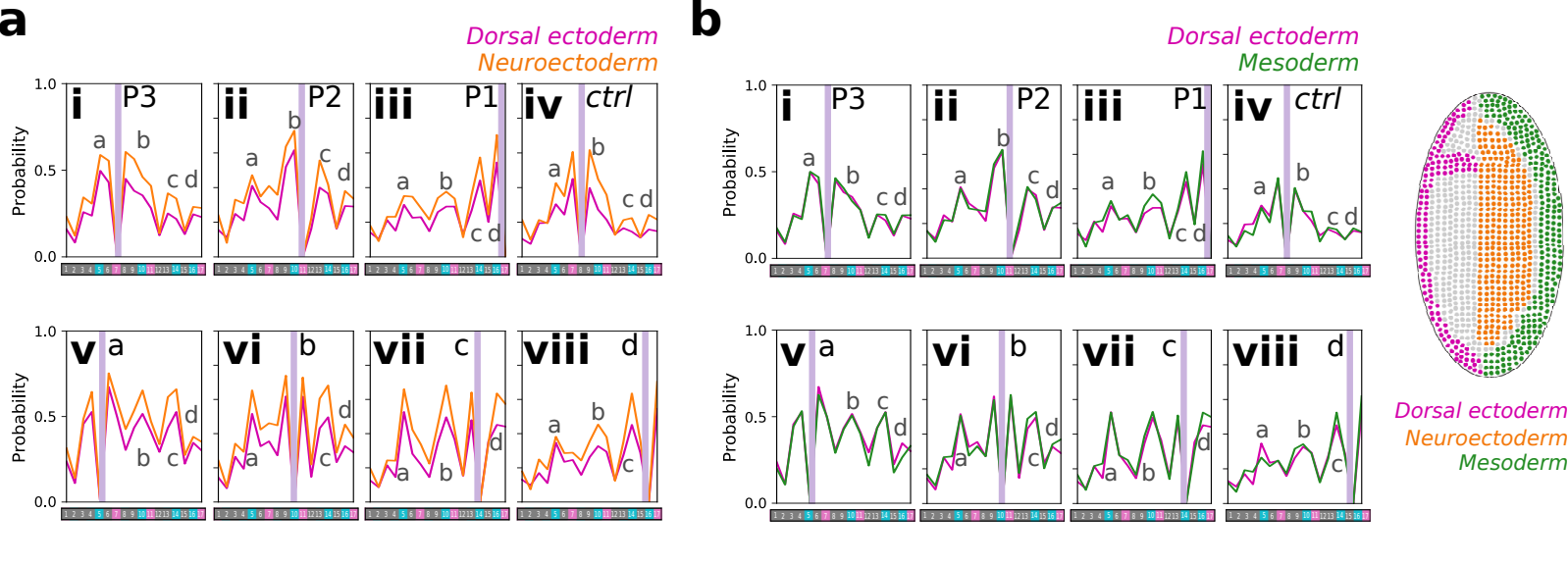

a
Low resolution Hi-M
N = 10927 | n = 9
nc14
nc11, nc12
N = 3782 | n = 7
doc-TAD
0.3
Probability
0.1
c
(Lott et al., 2011)
doc3
doc2
doc1
Normalized read count per gene
Nuclear cycle
nc10
nc11
nc12
nc13
nc14a
nc14b
nc14c
nc14d
b
i nc11
ii nc12
iii nc13
iv nc14
N = 1320 | n = 4
N = 2154 | n = 4
N = 7597 | n = 8
N = 37129 | n = 29
0.44
0.00
0.61
0.59
0.62
high
Probability
low
d
CRM
Probability
nc11
nc14
nc12
nc13
a
b
c
d
g
Promoter
P1
e
CRM
nc14
C
h
Promoter
P1
0.41
0.00
f
nc11
nc12
nc13
nc14

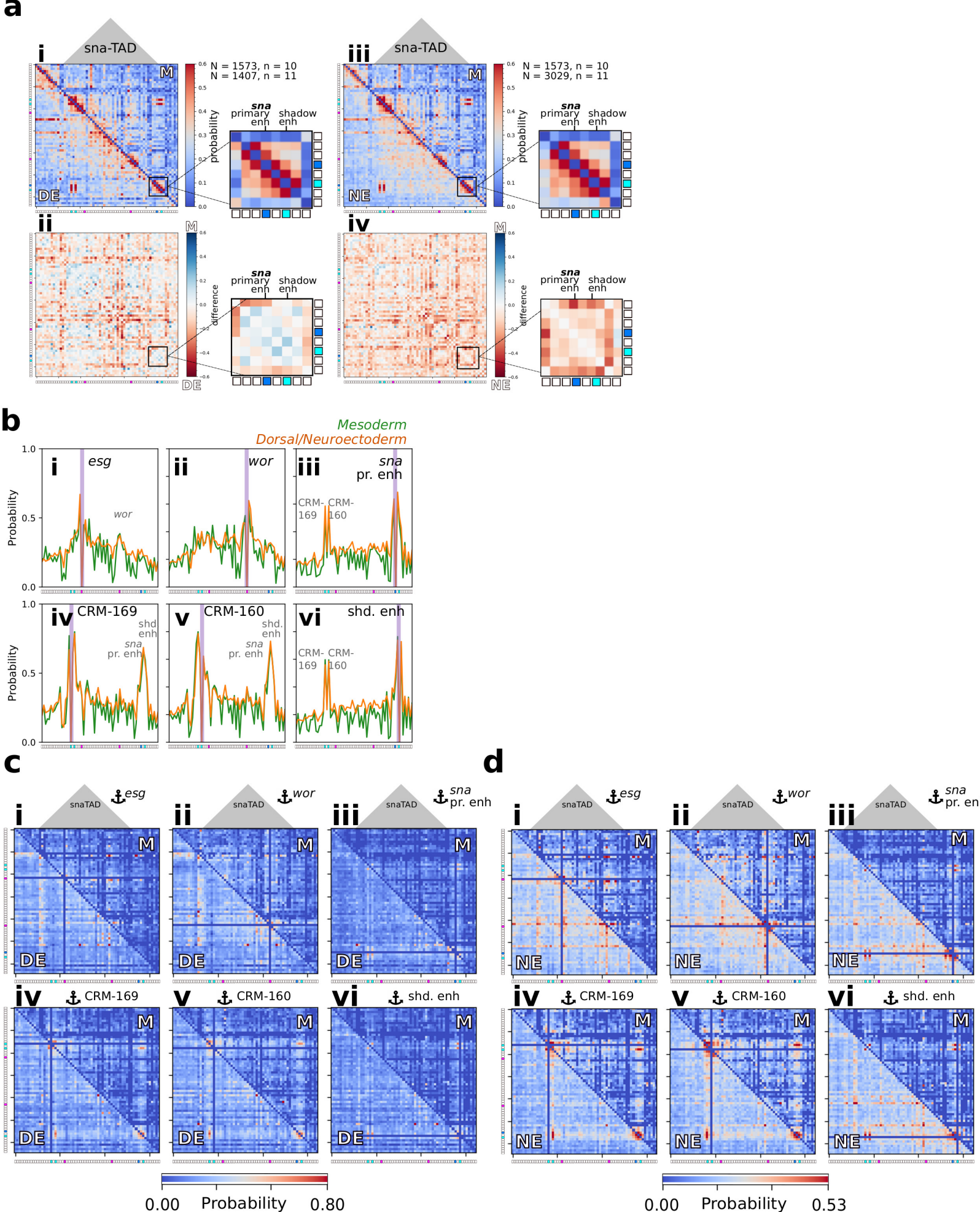
Figure 4

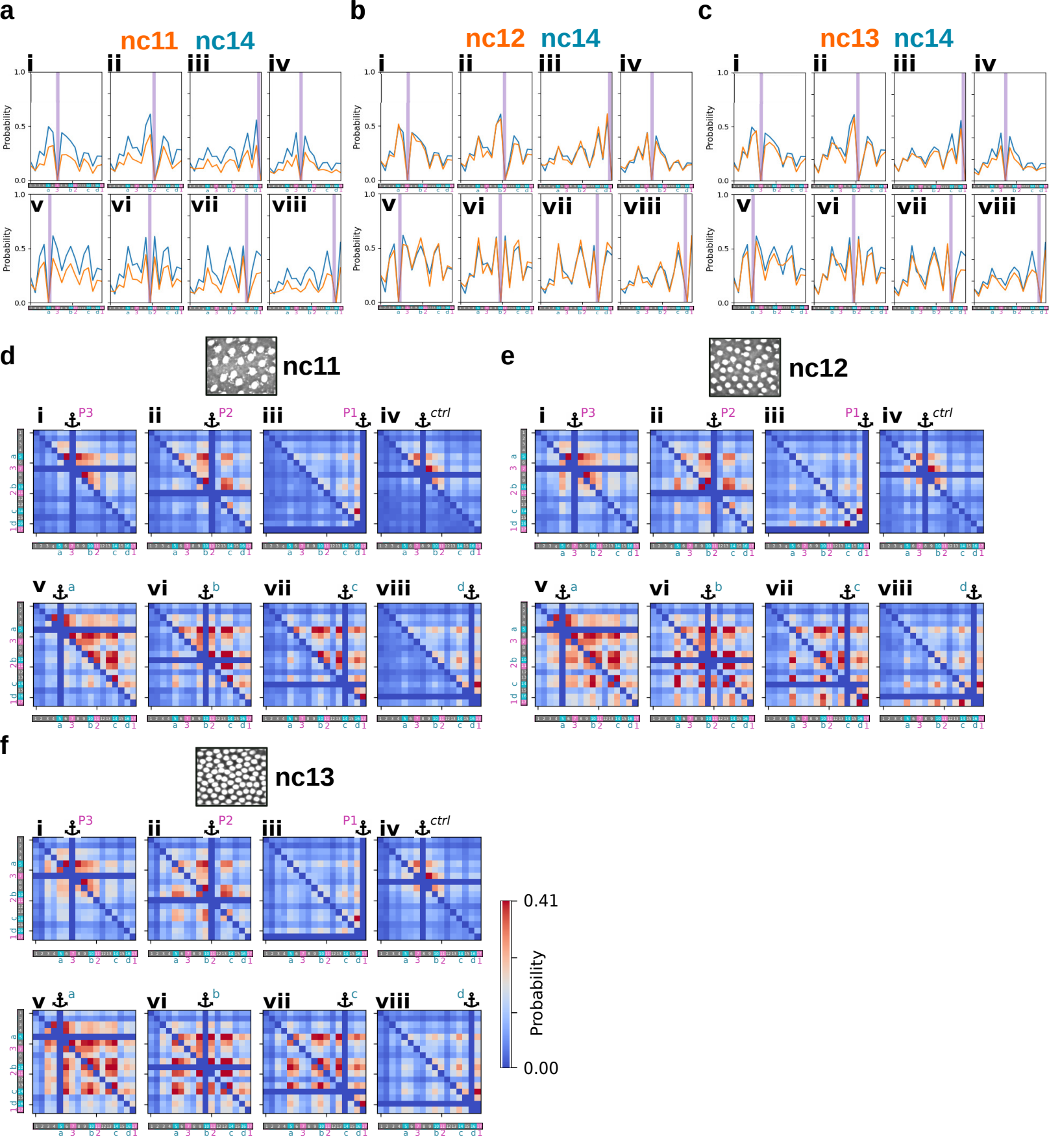

a
i
ii
Enhancers
Barcodes 10
17
Genes
CRMc
doc enhancer
DocF5
DocF5SacIIb
DocF5SacIIa
b
Doc3 Doc2
CRMc
enhancer
Doc1
Cas9
FRT
FRT
+Flippase
c
N = 21422 | n = 11
CRMc del
d
wt ctrl vs. CRMc deletion
Probability
P1
ctrl
e
P1
CRMc del
Probability
f
10 kb
wt nc14
-zld nc14
log2 ratio
-zld/wt
Gene
Barcodes 1
17
g
w-RNAi: N = 4603 | n = 4
zld-RNAi: N = 8432 | n = 8
w-RNAi
zld-RNAi
probability
h
w-RNAi vs. zld-RNAi
Probability
i
P1
w-RNAi
zld-RNAi
Probability
j
wt ctrl
zld-RNAi
ensemble
ensemble
Nucleus #474
k
nc14 ≤ 250 kb
Zelda enrichment
Zelda enrichment
log2(observed/expected)
l
nc14 zld-RNAi ≤ 250 kb
m
nc14 alpha amanitin ≤ 250 kb
n
nc12 ≤ 250 kb
o
nc 13 ≤ 250 kb

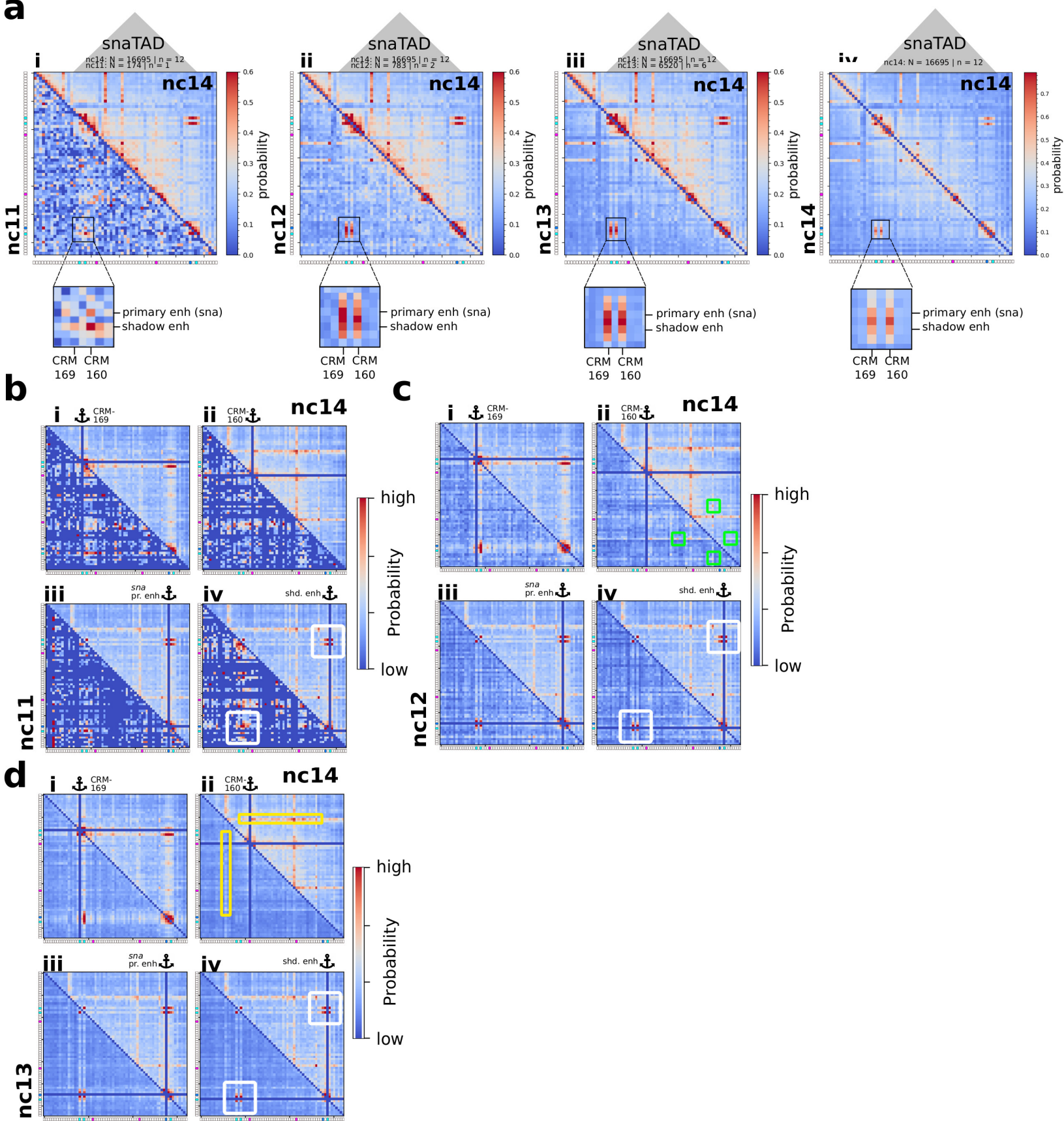

a
TAD 1
TAD 2 (doc-TAD)
165 kb
LowRes-HiM barcodes
HiRes-HiM barcodes
Genes
b
Barcodes (1-17)
CTCF
CP190
BEAF-32
su(Hw)
Mod(mdg4)
GAF
Zld
CBP
RAD21
CAP-H2
Chromator
Z4
TFIIIC
DREF
L3mbt
Genes
doc3
doc2
doc1
3L: 8.97 Mb
3L: 9.04 Mb
c
3µm
250nm
20µm
d
e
Efficiency
Barcode ID
Counts
Number of barcodes
f
max @ 0.3627
max @ 0.3552
max @ 0.5678
max @ 0.3577
max @ 0.3627
max @ 0.5553
max @ 0.2101
max @ 0.1726
max @ 0.3202
max @ 0.2076
max @ 0.1726
max @ 0.3077
max @ 0.2726
max @ 0.2626
max @ 0.4627
max @ 0.2726
max @ 0.2676
max @ 0.4477
g
| N = 37129 | n = 29
contact
| N = 37129 | n = 29
inverse PWD
inverse PWD, 1/ µm
h
Pearson corr coef subset vs full data
Fraction of cells from full dataset
i
counts
Rg, µm

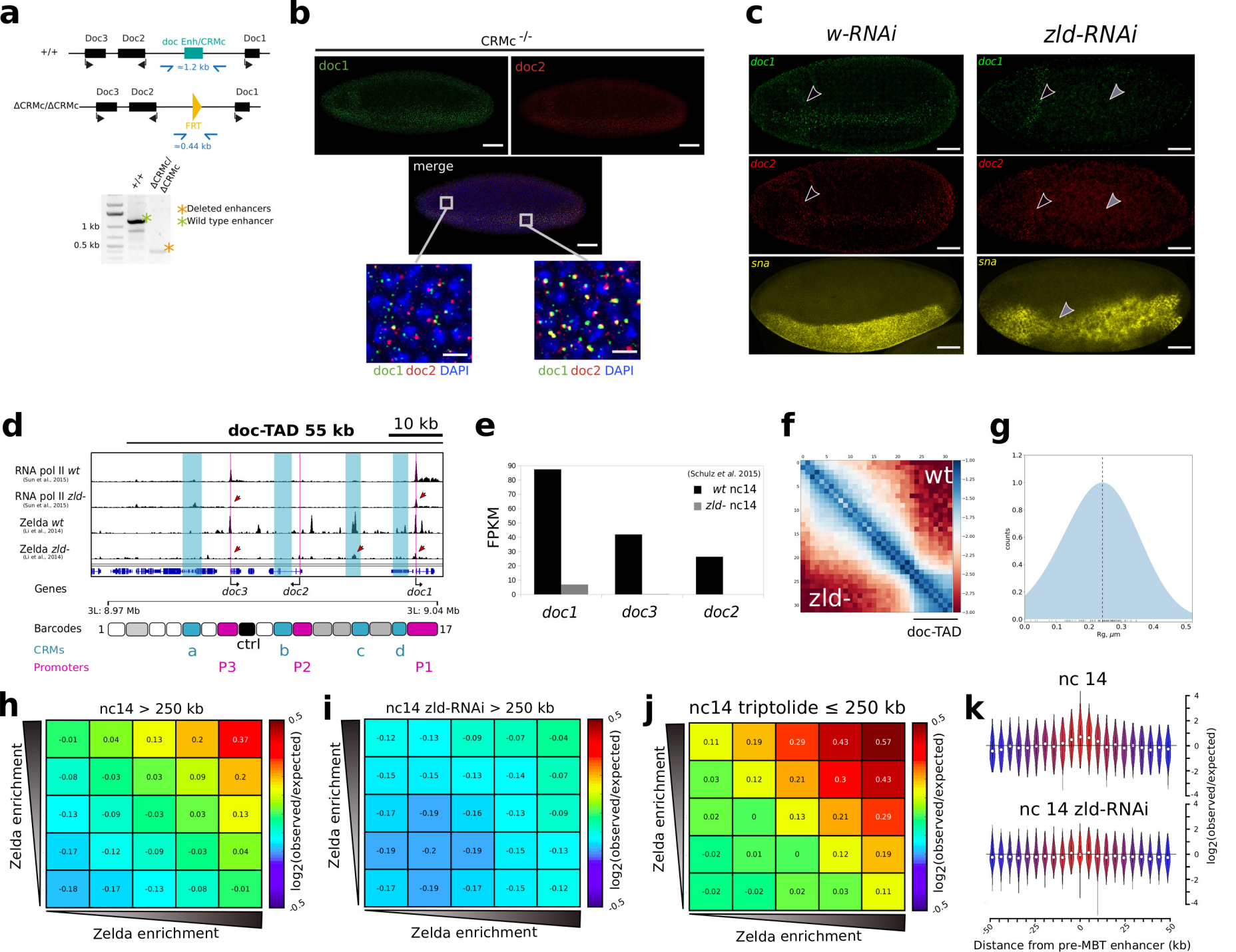

a
i
Doc1
Doc2
merge
DAPI Doc2 Doc1
ii
Doc3
Doc2
merge
DAPI Doc2 Doc3
b
Percentage of cells
doc1 transcription sites per cell
N=602, n=4
doc2 transcription sites per cell
c
N=1395, n=4
doc1
doc2
doc1&doc2
N=1243, n=3
doc2
doc3
doc2&doc3
d
DE
Subset (brightest)
Barcode #
Probability
e
Promoters
CRMs
P3
P2
P1
ctrl
f
Dorsal ectoderm
Subset (brightest)
P1
g
ATAC-seq
(Hannon et al., 2017)
Zelda
(Harrison et al., 2011)
RNA-seq
(Lott et al., 2011)
H3K4me1
(Koenecke et al. 2016)
Genes
nht
esg
sna
Tim17b2
2L: 15.30 Mb
2L: 15.34 Mb
2L: 15.47 Mb
2L: 15.49 Mb
Barcodes
sna-CRMs
CRM 169
CRM 160
primary enh
shadow enh
h
sna-TAD
N = 1573 | n = 12
55 kb
barcode #
probability
primary enh (sna)
shadow enh
CRM 169
CRM 160
esg
i
esg
wor
sna pr. enh
sna-TAD
Promoters
CRMs
CRM-169
CRM-160
shd. enh
0.57
Probability
0.00
j


a
Dorsal ectoderm
Neuroectoderm
i P3
ii P2
iii P1
iv ctrl
v a
vi b
vii c
viii d
Probability
b
Dorsal ectoderm
Mesoderm
Dorsal ectoderm
Neuroectoderm
Mesoderm


a
i
sna-TAD
M
DE
N = 1573, n = 10
N = 1407, n = 11
probability
sna
primary enh
shadow enh
iii
sna-TAD
M
NE
N = 1573, n = 10
N = 3029, n = 11
probability
sna
primary enh
shadow enh
ii
M
DE
difference
sna
primary enh
shadow enh
iv
M
NE
difference
sna
primary enh
shadow enh
b
Mesoderm
Dorsal/Neuroectoderm
i
esg
wor
ii
wor
iii
sna
pr. enh
CRM-169
CRM-160
Probability
iv
CRM-169
shd enh
sna
pr. enh
v
CRM-160
shd. enh
sna
pr. enh
vi
shd. enh
CRM-169
CRM-160
Probability
c
i
snaTAD
esg
M
DE
ii
snaTAD
wor
M
DE
iii
snaTAD
sna
pr. enh
M
DE
iv
CRM-169
M
DE
v
CRM-160
M
DE
vi
shd. enh
M
DE
0.00
Probability
0.80
d
i
snaTAD
esg
M
NE
ii
snaTAD
wor
M
NE
iii
snaTAD
sna
pr. enh
M
NE
iv
CRM-169
M
NE
v
CRM-160
M
NE
vi
shd. enh
M
NE
0.00
Probability
0.53


a
nc11
nc14
b
nc12
nc14
c
nc13
nc14
i
ii
iii
iv
v
vi
vii
viii
Probability
1.0
0.5
0.0
d
nc11
e
nc12
f
nc13
P3
P2
P1
ctrl
a
b
c
d
0.41
Probability
0.00


a
snaTAD
i
nc14: N = 16695 | n = 12
nc11: N = 174 | n = 1
nc14
nc11
probability
primary enh (sna)
shadow enh
CRM 169
CRM 160
ii
nc14: N = 16695 | n = 12
nc12: N = 783 | n = 2
nc12
iii
nc14: N = 16695 | n = 12
nc13: N = 6520 | n = 6
nc13
iv
nc14: N = 16695 | n = 12
nc14
b
nc14
CRM-169
CRM-160
sna pr. enh
shd. enh
high
Probability
low
nc11
c
nc12
d
nc13


a
Doc3
Doc2
doc Enh/CRMc
Doc1
+/+
≈1.2 kb
Doc3
Doc2
Doc1
ΔCRMc/ΔCRMc
FRT
≈0.44 kb
+/+
ΔCRMc/ΔCRMc
1 kb
0.5 kb
Deleted enhancers
Wild type enhancer
b
CRMc -/-
doc1
doc2
merge
doc1 doc2 DAPI
doc1 doc2 DAPI
c
w-RNAi
zld-RNAi
doc1
doc2
sna
d
doc-TAD 55 kb
10 kb
RNA pol II wt
(Sun et al., 2015)
RNA pol II zld-
(Sun et al., 2015)
Zelda wt
(Li et al., 2014)
Zelda zld-
(Li et al., 2014)
Genes
doc3
doc2
doc1
3L: 8.97 Mb
3L: 9.04 Mb
Barcodes
1
17
ctrl
CRMs
a
b
c
d
Promoters
P3
P2
P1
e
FPKM
(Schulz et al. 2015)
wt nc14
zld- nc14
doc1
doc3
doc2
f
wt
zld-
doc-TAD
g
counts
Rg, μm
h
nc14 > 250 kb
Zelda enrichment
Zelda enrichment
log2(observed/expected)
-0.01 0.04 0.13 0.2 0.37
-0.08 -0.03 0.03 0.09 0.2
-0.13 -0.09 -0.03 0.03 0.13
-0.17 -0.12 -0.09 -0.03 0.04
-0.18 -0.17 -0.13 -0.08 -0.01
i
nc14 zld-RNAi > 250 kb
Zelda enrichment
Zelda enrichment
log2(observed/expected)
-0.12 -0.13 -0.09 -0.07 -0.04
-0.15 -0.15 -0.13 -0.14 -0.07
-0.17 -0.19 -0.16 -0.13 -0.09
-0.19 -0.2 -0.19 -0.15 -0.13
-0.17 -0.19 -0.17 -0.15 -0.12
j
nc14 triptolide ≤ 250 kb
Zelda enrichment
Zelda enrichment
log2(observed/expected)
0.11 0.19 0.29 0.43 0.57
0.03 0.12 0.21 0.3 0.43
0.02 0 0.13 0.21 0.29
-0.02 0.01 0 0.12 0.19
-0.02 -0.02 0.02 0.03 0.11
k
nc 14
nc 14 zld-RNAi
log2(observed/expected)
Distance from pre-MBT enhancer (kb)